PHYSICS: A FOUNDATION FOR SCIENTISTS AND ENGINEERS



Michael Ragsdale
Fresno City College

LibreTexts™

Physics: A Foundation for Scientists and Engineers

Michael Ragsdale
Department of Astronomy, Engineering, and Physics
Fresno City College

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden borne from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and were supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 11/12/2026

TABLE OF CONTENTS

Licensing

Copyright

Preface

1: Measurement

1.1: What is Physics?

1.2: Measuring Standards

1.3: Scientific Notation and Significant Figures

1.4: Unit Conversions

1.5: Dimensional Analysis and Estimation

1.S: Measurement (Summary)

1.E: Measurement (Problems)

2: Motion in One Dimension

2.1: Position, Time, and Displacement

2.2: Speed and Velocity

2.3: Acceleration

2.4: Motion Graphs

2.5: Kinematics in One Dimension

2.6: Free Fall

2.S: Motion in One Dimension (Summary)

2.E: Motion in One Dimension (Problems)

3: Vectors

3.1: Scalars and Vectors

3.2: Components of a Vector

3.3: Unit Vectors

3.4: Vector Addition

3.5: Vector Multiplication

3.S: Vectors (Summary)

3.E: Vectors (Problems)

4: Motion in Two Dimensions

4.1: Kinematics in Two Dimensions

4.2: Projectile Motion

4.3: Uniform Circular Motion

4.4: Relative Motion

4.S: Motion in Two Dimensions (Summary)

4.E: Motion in Two Dimensions (Problems)

5: The Laws of Motion I

5.1: Newton's Laws of Motion

5.2: Types of Forces

5.3: Free-Body Diagrams

5.4: Equilibrium

5.5: Applying Newton's Laws

5.S: The Laws of Motion I (Summary)

5.E: The Laws of Motion I (Problems)

6: The Laws of Motion II

6.1: Friction

6.2: Resistive Forces

6.3: Uniform Circular Motion

6.S: The Laws of Motion II (Summary)

6.E: The Laws of Motion II (Problems)

7: Work and Power

7.1: Work Done by a Constant Force

7.2: Work Done by Gravity

7.3: Work Done by a General Force

7.4: Power

7.S: Work and Power (Summary)

7.E: Work and Power (Problems)

8: Conservation of Energy

8.1: Where Does the Work Go?

8.2: Kinetic Energy and the Work–Kinetic Energy Theorem

8.3: Potential Energy

8.4: Conservative and Nonconservative Forces

8.5: Conservation of Mechanical Energy

8.S: Conservation of Energy (Summary)

8.E: Conservation of Energy (Problems)

9: Linear Momentum and Center of Mass

9.1: Linear Momentum

9.2: Impulse

9.3: Conservation of Linear Momentum

9.4: Collisions

9.5: Center of Mass

9.S: Linear Momentum and Center of Mass (Summary)

9.E: Linear Momentum and Center of Mass (Problems)

10: Rotation I

10.1: Rotational Kinematics

10.2: Rolling Motion
10.3: Rotational Inertia
10.4: Torque and the Right-Hand Rule
10.S: Rotation I (Summary)
10.E: Rotation I (Problems)

11: Rotation II

11.1: Rotational Kinetic Energy
11.2: Rotational Work
11.3: Angular Momentum
11.4: Conservation of Angular Momentum
11.5: Precession
11.S: Rotation II (Summary)
11.E: Rotation II (Problems)

12: Static Equilibrium and Elasticity

12.1: Equilibrium
12.2: Center of Gravity and Stability
12.3: Elasticity
12.S: Static Equilibrium and Elasticity (Summary)
12.E: Static Equilibrium and Elasticity (Problems)

13: Universal Gravitation

13.1: Newton's Law of Universal Gravitation
13.2: Applications of Universal Gravitation
13.3: Gravitational Potential Energy
13.4: Kepler's Laws of Planetary Motion
13.S: Universal Gravitation (Summary)
13.E: Universal Gravitation (Problems)

14: Fluids

14.1: Fluids, Density, and Pressure
14.2: Pressure as a Function of Depth
14.3: Pascal's Principle
14.4: Buoyancy and Archimedes' Principle
14.5: Fluid Dynamics and Bernoulli's Principle
14.S: Fluids (Summary)
14.E: Fluids (Problems)

15: Electric Charges and Coulomb's Law

15.1: Electric Charge
15.2: Coulomb's Law
15.3: Discrete Charge Arrangements and the Superposition Principle
15.4: Electric Forces on Continuous Charge Distributions

15.S: Electric Charges and Coulomb's Law (Summary)
15.E: Electric Charges and Coulomb's Law (Problems)

16: Electric Fields

16.1: The Electric Field
16.2: The Electric Field of Point Charges
16.3: The Electric Field of Continuous Charge Distributions
16.4: A Point Charge in an Electric Field
16.5: A Dipole in an Electric Field
16.S: Electric Fields (Summary)
16.E: Electric Fields (Problems)

17: Gauss's Law

17.1: The Importance of Symmetry
17.2: Electric Flux
17.3: Gauss's Law
17.4: Applications of Gauss's Law
17.5: Conductors in Electrostatic Equilibrium
17.S: Gauss's Law (Summary)
17.E: Gauss's Law (Problems)

18: Electric Potential

18.1: Electric Potential
18.2: Equipotential Surfaces
18.3: Potential of a Discrete Charge Arrangement
18.4: Potential of a Continuous Charge Arrangement
18.5: Calculating the Field from the Potential
18.S: Electric Potential (Summary)
18.E: Electric Potential (Problems)

19: Capacitance

19.1: Capacitance
19.2: Energy Stored in an Electric Field
19.3: Dielectrics
19.4: Capacitors in Parallel and in Series
19.S: Capacitance (Summary)
19.E: Capacitance (Problems)

20: Current and Resistance

20.1: Electric Current
20.2: Resistance
20.3: Ohm's Law
20.4: Electrical Power
20.S: Current and Resistance (Summary)

20.E: Current and Resistance (Problems)

21: DC Circuits

21.1: Circuit Elements and Diagrams

21.2: Kirchhoff's Laws and the Basic Circuit

21.3: Multiple Resistors in a Circuit: Series vs. Parallel

21.4: Ideal vs. Real Batteries

21.5: RC Circuits

21.S: DC Circuits (Summary)

21.E: DC Circuits (Problems)

22: The Nature of Magnetic Fields

22.1: Magnetic Fields

22.2: Motion of a Charged Particle in Magnetic Fields

22.3: Applications of Magnetic Fields

22.S: The Nature of Magnetic Fields (Summary)

22.E: The Nature of Magnetic Fields (Problems)

23: Magnetic Fields and Currents

23.1: Magnetic Field Due to a Current

23.2: Force Between Two Parallel Currents

23.3: Magnetic Dipoles

23.4: Ampère's Law

23.5: Solenoids and Toroids

23.6: Gauss's Law in Magnetism

23.7: Magnetic Properties of Matter

23.S: Magnetic Fields and Currents (Summary)

23.E: Magnetic Fields and Currents (Problems)

24: Faraday's Law

24.1: Magnetic Flux

24.2: Faraday's Law of Induction

24.3: Motional EMF

24.4: Generators and Motors

24.S: Faraday's Law (Summary)

24.E: Faraday's Law (Problems)

25: Inductance

25.1: Inductors

25.2: Energy in a Magnetic Field

25.3: RL Circuits

25.4: Transformers

25.S: Inductance (Summary)

25.E: Inductance (Problems)

26: AC Circuits

26.1: LC Oscillations

26.2: Resistors in an AC Circuit

26.3: Capacitors in an AC Circuit

26.4: Inductors in an AC Circuit

26.5: Series RLC Circuits

26.6: Power in AC Circuits

26.S: AC Circuits (Summary)

26.E: AC Circuits (Problems)

27: Electromagnetic Waves

27.1: Displacement Current

27.2: Maxwell's Equations

27.3: Propagation of Electromagnetic Waves

27.4: Energy Carried by Electromagnetic Waves

27.5: Momentum and Radiation Pressure of Electromagnetic Waves

27.6: The Electromagnetic Spectrum

27.S: Electromagnetic Waves (Summary)

27.E: Electromagnetic Waves (Problems)

28: Oscillations

28.1: Simple Harmonic Motion

28.2: Hooke's Law

28.3: Mass–Spring System

28.4: Energy in Simple Harmonic Motion

28.5: The Simple Pendulum

28.6: Torsion Pendulum

28.7: Physical Pendulum

28.8: Damped Oscillations

28.9: Forced Oscillations

28.S: Oscillations (Summary)

28.E: Oscillations (Problems)

29: Waves I

29.1: Wave Motion

29.2: The Wave Equation

29.3: Speed of a Wave

29.4: Energy of a Wave

29.5: Phasors

29.6: Standing Waves and Resonance

29.S: Waves I (Summary)

29.E: Waves I (Problems)

30: Waves II

30.1: Sound Waves

30.2: Speed of Sound

30.3: Wave Interference

30.4: Intensity and Decibels

30.5: Beats

30.6: The Doppler Effect

30.S: Waves II (Summary)

30.E: Waves II (Problems)

31: Introduction to Optics

31.1: Light as a Wave

31.2: Ray Optics Approximation

31.3: Polarization

31.4: Reflection and Refraction

31.5: Total Internal Reflection

31.6: Polarization by Reflection

31.S: Introduction to Optics (Summary)

31.E: Introduction to Optics (Problems)

32: Images

32.1: Images and Plane Mirrors

32.2: Spherical Mirrors

32.3: Thin Lenses

32.4: Optical Systems

32.5: Optical Instruments

32.S: Images (Summary)

32.E: Images (Problems)

33: Interference

33.1: Young's Double-Slit Experiment

33.2: Interference and Double-Slit Intensity

33.3: Interference from Thin Films

33.4: Interferometry

33.S: Interference (Summary)

33.E: Interference (Problems)

34: Diffraction

34.1: Single-Slit Diffraction

34.2: Diffraction by a Circular Aperture

34.3: Diffraction by a Double Slit

34.4: Diffraction Gratings

34.5: Diffraction of X-Rays by Crystals

34.S: Diffraction (Summary)
34.E: Diffraction (Problems)

35: Relativity

35.1: History and Motivation for Special Relativity
35.2: Consequences of Special Relativity
35.3: Lorentz Transformations
35.4: Doppler Effect for Light
35.5: Momentum and Energy
35.6: General Relativity
35.S: Relativity (Summary)
35.E: Relativity (Problems)

36: Temperature and Kinetic Theory of Gases

36.1: Temperature
36.2: Thermal Expansion
36.3: Ideal Gas Law
36.4: Kinetic Theory of Gases
36.S: Temperature and Kinetic Theory of Gases (Summary)
36.E: Temperature and Kinetic Theory of Gases (Problems)

37: Heat and the First Law of Thermodynamics

37.1: Heat
37.2: The First Law of Thermodynamics
37.3: Thermodynamic Processes
37.4: Heat Transfer Mechanisms
37.S: Heat and the First Law of Thermodynamics (Summary)
37.E: Heat and the First Law of Thermodynamics (Problems)

38: Entropy and the Second Law of Thermodynamics

38.1: Entropy
38.2: The Second Law of Thermodynamics
38.3: Heat Engines
38.S: Entropy and the Second Law of Thermodynamics (Summary)
38.E: Entropy and the Second Law of Thermodynamics (Problems)

39: Introduction to Quantum Mechanics

39.1: Blackbody Radiation
39.2: Photoelectric Effect
39.3: Photons
39.4: Compton Shift
39.5: de Broglie Wavelength
39.6: Heisenberg Uncertainty Principle
39.S: Introduction to Quantum Mechanics (Summary)

39.E: Introduction to Quantum Mechanics (Problems)

40: A Deeper Look at Quantum Mechanics

40.1: Schrödinger's Equation and Wave Functions

40.2: Energies of Trapped Electrons

40.3: Infinite Wells

40.4: Finite Wells

40.5: Two- and Three-Dimensional Wells

40.S: A Deeper Look at Quantum Mechanics (Summary)

40.E: A Deeper Look at Quantum Mechanics (Problems)

41: Atomic Physics

41.1: Models of the Atom

41.2: Energies of a Trapped Electron in an Atom

41.3: The Hydrogen Atom

41.4: Atomic Spectra of Gases

41.5: The Quantum Model of the Hydrogen Atom

41.6: The Pauli Exclusion Principle

41.7: Lasers

41.S: Atomic Physics (Summary)

41.E: Atomic Physics (Problems)

42: Bonding and Electrical Conduction in Solids

42.1: Molecular Bonds

42.2: Free-Electron Theory of Metals

42.3: Electrical Conduction in Metals and Insulators

42.4: Semiconductors

42.S: Bonding and Electrical Conduction in Solids (Summary)

42.E: Bonding and Electrical Conduction in Solids (Problems)

43: Nuclear Physics

43.1: Properties of Nuclei

43.2: Nuclear Binding Energy

43.3: Radioactivity

43.4: Nuclear Reactions

43.S: Nuclear Physics (Summary)

43.E: Nuclear Physics (Problems)

44: Particle Physics and Cosmology

44.1: Elementary Particles

44.2: Basic Forces and Messenger Particles

44.3: The Standard Model

44.4: Cosmology

44.S: Particle Physics and Cosmology (Summary)

44.E: Particle Physics and Cosmology (Problems)

A: Physical Constants and Astronomical Data

B: SI Units, Prefixes, and Conversion Factors

C: Basic Algebra and Trigonometry for Physics Students

C.1: Solving an Equation for One Unknown

C.2: Quadratic Equations

C.3: Solving for Two Unknowns

C.4: Solving for Three Unknowns

C.5: Right-Triangle Trigonometry

C.6: Breaking a Vector into Components

C.7: Finding the Hypotenuse from the Components

C.8: Sinusoids and Phasors

C.S: Basic Algebra and Trigonometry for Physics Students (Summary)

C.E: Basic Algebra and Trigonometry for Physics Students (Practice Problems)

D: Motivation of Calculus for Physics Students

D.1: The Derivative: Slope as a Rate of Change

D.2: The Integral: Adding Up Small Pieces

D.3: Why the Two Ideas Undo Each Other

D.4: Integration in Physics: Slice, Label, Add

D.5: Checking a Calculus Result

D.S: Motivation of Calculus for Physics Students (Summary)

D.E: Motivation of Calculus for Physics Students (Practice Problems)

Index

Detailed Licensing

This page titled Physics: A Foundation for Scientists and Engineers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

Copyright

Physics: A Foundation for Scientists and Engineers, First Edition.

Copyright © 2026 Michael Ragsdale.

This book is an open educational resource. It is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

You are free to share (copy and redistribute the material in any medium or format) and to adapt (remix, transform, and build upon the material) for any purpose, including commercially, provided that you give appropriate credit, provide a link to the license, and indicate whether changes were made.

The full text of the license is available at the Creative Commons website.

The complete LaTeX source for this book is available so that it may be corrected, adapted, and rebuilt.

The author has made every effort to ensure the accuracy of the information herein. However, the information contained in this book is provided without warranty, either express or implied.

This page titled Copyright is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

Preface

This textbook is intended for a first-year, calculus-based introductory physics course for students majoring in science and engineering. It assumes familiarity with single-variable differential and integral calculus and develops physical reasoning alongside mathematical technique.

Each chapter follows the same structure: a conceptual introduction, the formal development of the physics, worked examples that demonstrate problem-solving strategies, conceptual checkpoints to test understanding, a chapter summary, and a graded set of end-of-chapter problems.

An open educational resource

This book is an **open educational resource** (OER). It is free to read, free to download, and free to keep: there is no access code, no expiring rental, and no new edition issued for the purpose of making the previous one worthless. Students should never have to pay for access to the material their course requires, and the cost of a commercial physics text is a real barrier to real students.

Being open means more than being free of charge. The license below grants you the right to *use* this book, to *redistribute* it, and to *adapt* it. Instructors are encouraged to do all three. If your course covers chapters in a different order, drop the ones you do not need. If your students would be better served by different examples (ones drawn from your region, your industry, or your laboratory) rewrite them. If you find an explanation that does not land with your students, replace it with one that does. You do not need to ask permission, and you do not need to explain yourself.

The complete LaTeX source is distributed along with the book so that adaptation is genuinely practical rather than merely permitted. A PDF alone can be read; source can be edited, corrected, extended, and rebuilt.

License

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). In plain terms, you are free to:

- **Share**: copy and redistribute the material in any medium or format;
- **Adapt**: remix, transform, and build upon the material for any purpose, including commercially.

These freedoms are subject to one condition:

- **Attribution**: you must give appropriate credit, provide a link to the license, and indicate if any changes were made. You may do so in any reasonable manner, but not in any way that suggests the author endorses you or your use of the work.

The full license is available at the Creative Commons website.

This is the most permissive Creative Commons license. It lets the book be combined with material under almost any other open license, including other CC BY and CC BY-SA resources, and reused in courseware, homework systems and adaptations of every kind. The only requirement is credit.

Attribution

If you adapt or redistribute this book, an attribution of the following form is sufficient:

> *Physics: A Foundation for Scientists and Engineers* by Michael Ragsdale, used under CC BY 4.0.

If you have modified the work, say so, for example, by preceding the attribution with “Adapted from” and briefly describing what you changed.

Corrections and contributions

Errors in a physics textbook are not merely embarrassing; they cost students hours. If you find one (a wrong sign, a bad numerical answer, a figure that contradicts its caption, an explanation that misleads) please report it. Corrections from instructors and students alike are welcome, and they are the most valuable contribution anyone can make to a book like this one.

Michael Ragsdale
Department of Astronomy, Engineering, and Physics
Fresno City College
August 2026

This page titled Preface is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1: MEASUREMENT

A student's dormitory room measures $3.1\ \mathrm{m}$ by $3.8\ \mathrm{m}$ and has a ceiling $2.6\ \mathrm{m}$ high. After completing their physics course, the student decides to celebrate by wallpapering every wall with the pages from their physics textbook: 960 pages of text printed on 480 sheets, each $21\ \mathrm{cm} \times 28\ \mathrm{cm}$. The pages cover even the door and window. Is this possible?

This problem looks like arithmetic, but it captures something essential about physics: we cannot answer it without first deciding what to measure, what units to use, and how precisely we can compute the answer. *Measurement* is the foundation on which every later chapter of this book rests. Before we can describe motion, force, or energy, we have to agree on how to quantify length, mass, and time, and on how to manipulate numbers that span dozens of orders of magnitude, from the lower limit on the lifetime of a proton (more than $10^{41}\ \mathrm{s}$) to the duration of the shortest laser pulse yet produced ($\sim 10^{-17}\ \mathrm{s}$).

Learning Objectives

- Describe what physics is and how a physics problem differs from a pure math problem.
- Identify the SI base units of length, mass, and time, and express derived units in terms of them.
- Write numbers in scientific notation and use SI prefixes to compress large and small values.
- Convert quantities between units, including compound units and units raised to powers.
- Use dimensional analysis and significant figures to check and report calculations.
- Make order-of-magnitude estimates and judge when an estimate is good enough.

(You will return to the dorm-room problem in Problem 1.19 at the end of this chapter, once you have the tools to answer it.)

1.1: What is Physics?

1.2: Measuring Standards

1.3: Scientific Notation and Significant Figures

1.4: Unit Conversions

1.5: Dimensional Analysis and Estimation

1.S: Measurement (Summary)

1.E: Measurement (Problems)

This page titled 1: Measurement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.1: What is Physics?

Physics is the natural science concerned with the study of matter, energy, space, and time, and the relationships among them. It is one of the most fundamental branches of science: where chemistry asks about atoms and molecules, biology about living organisms, and geology about the Earth, physics asks about the rules those domains all obey. Its goal is to identify the smallest set of principles (often expressible as mathematical equations) from which the behavior of the universe at every scale can be derived.

More broadly, **science** is the systematic study of the structure and behavior of the physical and natural world through observation, experimentation, and the testing of theories against empirical evidence. What distinguishes a scientific claim from any other kind of claim is not the use of mathematics or laboratory equipment, but a commitment to letting evidence decide.

Math problems vs. physics problems

A math problem typically begins with a set of given equations or symbols and asks you to manipulate them according to formal rules to reach a definite answer. The setup is unambiguous; the work is in the algebra.

A physics problem usually begins with a description of a real-world situation: a falling ball, a sliding block, a circuit with three resistors. Most of the work is the modeling: deciding which physical principles apply, which quantities matter, what idealizations are safe to make, and only *then* translating the situation into equations. The algebra at the end is often the easiest step.

A strategy for solving physics problems

If the modeling is the hard part, it helps to have a procedure for doing it. The following five steps are used throughout this book, and the worked examples in every chapter follow them whether or not the headings are printed.

1. **Draw the situation and label it.** Sketch what is happening, even when the picture seems obvious. Mark the objects, the directions of motion if there is any, and your choice of coordinate axes. The sketch is where you *decide* which way is positive, and many sign errors are made by students who never wrote that choice down. This is where you get to choose. You have control. Don't give that up. List what you are given and what you are asked for, assigning a symbol to each. Again, you have control over the choice of symbol. Just because this textbook uses v_0 for the initial velocity does not mean you can't use v_i if that works better for your brain. We all think differently, and that's a beautiful thing.
2. **Identify the physics.** Ask which principle governs the situation before reaching for any formula. Is the quantity in question conserved? Is the net force zero? Is the acceleration constant? Choosing a principle is a physical judgment; the equation that follows is merely its accounting. This is also the moment to state your **idealizations** (frictionless, massless, negligible air resistance) so you know exactly what you have assumed and what would change if the assumption failed.
3. **Solve symbolically.** Do the math (this is where your algebra, trigonometry, and calculus come into play) in symbols and substitute numbers only at the end. A symbolic result shows which quantities actually matter, is far easier to check, and can be reused for any set of values.
4. **Substitute and evaluate.** Put in the numbers with their units, carry the units through the arithmetic, and round the final answer to the appropriate number of significant figures.
5. **Check the answer.** Never stop at a boxed number. Do the units come out right? Is the value plausible (is your car traveling at a reasonable speed, or at three times the speed of light)? Does the sign point the way your sketch says it should? If the answer to even one of these questions is no, something is probably wrong.

Note: Why symbols before numbers

Substituting numbers early is the single most common way to make a problem harder than it is. Once a quantity becomes $9.8\ \mathrm{m/s^2}$ it stops being recognizable as g, and an expression that would have simplified is now a string of decimals that cannot. Worse, an early numerical slip propagates silently to the end, whereas a symbolic result can be inspected: if the mass fails to cancel where you expected it to, you know something is wrong before you have wasted any arithmetic.

The final step deserves more weight than it usually gets. Two checks are cheap and catch a surprising number of errors. The first

is dimensional: an equation whose two sides have different dimensions is certainly wrong, and this test costs seconds (Section 1.5). The second is to examine **limiting cases**. Push a symbolic result to an extreme where you already know the answer. If a formula for two masses connected by a cord does not reduce to free fall when one mass goes to zero, the formula is wrong, and you have found out without checking a single line of algebra.

Finally, a word about being stuck, which is a perfectly normal thing and not a verdict on your ability. Everyone struggles with this when they start. It's okay. Being stuck almost always means step 1 or step 2 is incomplete rather than that you have forgotten an equation. Return to the sketch, write down every quantity you know, and ask what physical principle connects them. Searching for a formula that contains the right letters is the least reliable way to solve a physics problem, and it is the habit this book is written to replace. You can do it!

This page titled 1.1: What is Physics? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.2: Measuring Standards

Every measurement is fundamentally a *comparison* between the object being measured and an agreed-upon **standard**, called a **unit**. To say a hallway is “three meters long” is to say that its length contains three copies of the standard meter. Without a shared standard, the statement is meaningless: three of *whose* meters?

Although many measurement systems have existed historically, the modern scientific standard is the **International System of Units**, abbreviated **SI** (from the French *Système International d’Unités*). Throughout this book we use SI units exclusively; U.S. customary units (feet, pounds, gallons) will appear only as quantities to be *converted into* SI. Be careful with the word “gallon”: the U.S. gallon and the imperial gallon still used in parts of the Commonwealth differ by about 20%. Every gallon in this book is a U.S. gallon.

The SI base units

The SI system is built on seven **base units**, of which three are used pervasively in introductory mechanics:

The three SI base units used most in mechanics

Quantity	SI unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s

The remaining four base units (electric current: ampere, A; temperature: kelvin, K; amount of substance: mole, mol; and luminous intensity: candela, cd) will appear in later chapters as needed.

Definitions of the base units

Since 2019, all SI base units have been defined in terms of fixed numerical values of fundamental physical constants of nature. This represents a major shift from earlier definitions, which depended on physical artifacts (like a particular metal bar held in a vault) or specific material properties that could be lost, damaged, or drift over time. The modern definitions are reproducible in any properly equipped laboratory anywhere in the universe.

The second (s). The second is defined by fixing the numerical value of a particular atomic transition frequency in cesium-133. In the ground state of a cesium atom, the outermost electron can sit in either of two slightly different energy levels, depending on how its magnetic orientation lines up with that of the nucleus: this pair of closely spaced levels is called a *hyperfine* splitting. Flipping between them absorbs or emits microwave radiation at one fixed frequency. (“Unperturbed” simply means the atom is shielded from stray electric and magnetic fields that would shift this frequency slightly.) That frequency is defined to be exactly

$$\Delta\nu_{\mathrm{Cs}} = 9\,192\,631\,770\,\mathrm{Hz}. \tag{1.2.1}$$

One second is therefore the duration of exactly $9\ 192\ 631\ 770$ cycles of this radiation. Atomic clocks built around this transition are the most precise timekeepers ever constructed, drifting by less than a second over hundreds of millions of years.

The meter (m). The meter is defined by fixing the numerical value of the speed of light in vacuum to be exactly

$$c = 299\,792\,458\,\mathrm{m/s}. \tag{1.2.2}$$

One meter is the distance light travels in vacuum during a time interval of $1/299\ 792\ 458$ of a second. Because the second is already defined via cesium, this fixes the meter as well. Note the consequence: we no longer *measure* the speed of light; we *define* it. Any future improvement in our ability to compare light travel times does not sharpen our value of c, which is now exact by definition; instead it lets a laboratory build a better physical embodiment of the meter itself. (Metrologists call such a physical embodiment a *realization* of the unit.)

The kilogram (kg). The kilogram is defined by fixing the numerical value of the Planck constant to be exactly

$$h = 6.626\,070\,15 \times 10^{-34}\ \mathrm{J \cdot s}. \tag{1.2.3}$$

Because $1\ \mathrm{J \cdot s} = 1\ \mathrm{kg \cdot m^2/s}$, and the meter and second are already defined, this fixes the kilogram. This 2019 redefinition retired the last physical-artifact standard: the International Prototype of the Kilogram, a platinum–iridium cylinder held in a vault near Paris that had served as the world's mass standard since 1889.

Note: Where the meter came from

The meter began in 1791, when the French Academy of Sciences defined it as one ten-millionth of the distance from the North Pole to the equator, measured along the meridian through Paris. Two astronomers, Jean-Baptiste Delambre and Pierre Méchain, spent 1792 to 1798 surveying the stretch of that meridian between Dunkirk and Barcelona, and in 1799 the result was cast as a platinum bar. The survey was a remarkable achievement, but the Earth's shape was not known well enough at the time, and the bar came out about $0.2\ \mathrm{mm}$ short of its intended length. That is why the Earth's circumference is roughly, but not exactly, $40\ 000\ \mathrm{km}$.

In 1889 the standard became the International Prototype Meter, a platinum–iridium bar whose length was the distance between two fine lines engraved on it, measured at the melting point of ice. It was an artifact standard, with the same weaknesses as the kilogram cylinder. In 1960 the meter was redefined as $1\ 650\ 763.73$ wavelengths of an orange-red spectral line of krypton-86: the first atomic definition, which any laboratory could reproduce. In 1983 it took its present form, tied to the speed of light.

Notice what stayed fixed through every change. Each new definition was chosen to agree with the old meter as closely as it could then be measured, so the length of a meter never jumped; only the precision with which it could be realized improved. The other base units followed the same path. The second began as $1/86\ 400$ of a mean solar day, but the Earth's rotation is not perfectly steady (it is gradually slowing), so in 1967 the second was tied to cesium. The kilogram began as the mass of one liter of water, became a platinum cylinder in 1799 and the platinum–iridium prototype in 1889, and was tied to the Planck constant in 2019.

Derived units

Most physical quantities you will encounter (velocity, force, energy, pressure) are not base quantities. They are **derived**: each has a unit built out of products and ratios of the base units. For example, the unit of force is the **newton**, defined as

$$1\,\mathrm{N} = 1\,\frac{\mathrm{kg \cdot m}}{\mathrm{s^2}}. \tag{1.1}$$

A newton is the force required to accelerate a $1\ \mathrm{kg}$ mass at $1\ \mathrm{m/s^2}$. We will derive this expression from Newton's second law in Chapter 5.

This page titled 1.2: Measuring Standards is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.3: Scientific Notation and Significant Figures

Physics deals with quantities spanning an extraordinary range of magnitudes. The age of the universe is roughly 4×10^{17} seconds; the Planck time (the scale below which our current theories of space and time are expected to break down, so that we do not know how to describe shorter intervals) is about 1×10^{-43} seconds. Writing such numbers in ordinary decimal form is impractical and error-prone. We therefore use **scientific notation**, expressing a quantity as a single nonzero digit before the decimal point, multiplied by a power of ten:

$$N = a \times 10^{n}, \qquad 1 \le |a| < 10, \quad n \in \mathbb{Z}. \tag{1.2}$$

For example, $530\ 000\ 000\ \text{m} = 5.3 \times 10^{8}\ \text{m}$, and $0.000\ 000\ 042\ \text{s} = 4.2 \times 10^{-8}\ \text{s}$. Some characteristic values:

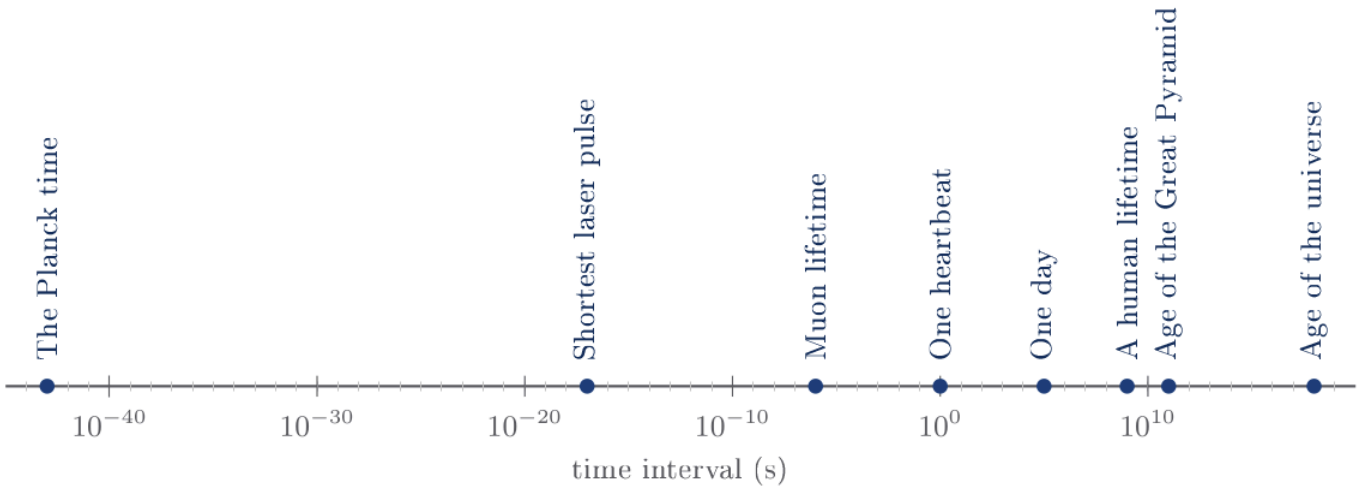


Figure 1.1. The range of time intervals in physics, on a logarithmic scale. Each labeled decade is ten times the one before it, so equal distances along the axis represent equal *factors*, not equal amounts.

•Figure description

A horizontal logarithmic axis of time intervals in seconds, running from ten to the minus forty-four seconds on the left to ten to the twentieth on the right, with major ticks labeled every ten decades. Eight events are marked along it: the Planck time at ten to the minus forty-three, the shortest laser pulse at ten to the minus seventeen, the muon lifetime at ten to the minus six, one heartbeat at ten to the zero, one day at ten to the fifth, a human lifetime at ten to the ninth, the age of the Great Pyramid at ten to the eleventh, and the age of the universe at ten to the eighteenth. Equal distances along the axis represent equal factors of ten, not equal amounts of time.

Some time intervals in physics, in seconds

Time interval (s)		Time interval (s)	
Age of the universe	4×10^{17}	Time between heartbeats	8×10^{-1}
Age of the Great Pyramid	1×10^{11}	Lifetime of the muon	2×10^{-6}
Human life expectancy	2×10^{9}	Shortest laser pulse made	4×10^{-17}
Length of a day	9×10^{4}	The Planck time	1×10^{-43}

The bottom of that range is a moving target. Laser pulses lasting only a few tens of *attoseconds* (10^{-17} s) can now be produced (work recognized by the 2023 Nobel Prize in Physics), and the record continues to fall.

A table makes the numbers available, but it does not make them *felt*. Figure 1.1 plots the same quantities on a logarithmic axis, where each step to the right multiplies the interval by ten. Sixty decades separate the shortest interval we can speak about from the age of the universe, and every physical process you will ever study lives somewhere on this line.

SI prefixes

A second tool for handling extreme magnitudes is the use of **SI prefixes**, which scale a base unit by a power of ten:

SI prefixes

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{30}	quetta-	Q	10^{-1}	deci-	d
10^{27}	ronna-	R	10^{-2}	centi-	c
10^{24}	yotta-	Y	10^{-3}	milli-	m
10^{21}	zetta-	Z	10^{-6}	micro-	μ
10^{18}	exa-	E	10^{-9}	nano-	n
10^{15}	peta-	P	10^{-12}	pico-	p
10^{12}	tera-	T	10^{-15}	femto-	f
10^{9}	giga-	G	10^{-18}	atto-	a
10^{6}	mega-	M	10^{-21}	zepto-	z
10^{3}	kilo-	k	10^{-24}	yocto-	y
10^{2}	hecto-	h	10^{-27}	ronto-	r
10^{1}	deka-	da	10^{-30}	quecto-	q

The four outermost entries (ronna, quetta, ronto, and quecto) were added to the SI only in 2022, largely because global data storage had begun to outgrow the old ceiling of yotta. The system is a living standard, not a finished one.

For example, $530\ 000\ 000\ \text{m} = 5.3 \times 10^{8}\ \text{m} = 530 \times 10^{6}\ \text{m} = 530\ \text{Mm}$ ("530 megameters"). The prefixes you will encounter most often in this course are *kilo-*, *centi-*, *milli-*, *micro-*, and *nano-*; the others appear in specialized contexts.

Significant figures

When you report a measured or computed quantity, the number of digits you write should reflect how precisely you actually know it. The digits you can defend are called **significant figures** (or *sig figs*). A measurement of $1.40\ \text{m}$ implies precision to the nearest centimeter (three significant figures); writing it as $1.4\ \text{m}$ implies precision only to the nearest decimeter (two significant figures), and $1.400\ \text{m}$ to the nearest millimeter (four).

Counting significant figures. Three rules settle nearly every case:

- Every nonzero digit is significant. (4.51 has three.)
- A zero *between* nonzero digits is significant. (4.05 has three; 1.0002 has five.)
- A *leading* zero is never significant: it only locates the decimal point (0.0027 has two, not four). A *trailing* zero is significant only when the number contains a decimal point (4.500 has four; 0.250 has three).

That last rule leaves a genuine ambiguity. Written as $1200\ \text{s}$, a value could mean two, three, or four significant figures, and nothing on the page tells you which. Scientific notation removes the doubt: 1.2×10^{3}, 1.20×10^{3}, and 1.200×10^{3} say two, three, and four unambiguously. Whenever trailing zeros carry meaning, write the number in scientific notation.

Exact numbers impose no limit. The 2 in the circumference $2\pi r$ is a fact of geometry, not a measurement. Defined conversion factors are exact by agreement: $1\ \text{in} = 2.54\ \text{cm}$ and $1\ \text{min} = 60\ \text{s}$ hold to as many digits as you care to write. Exact numbers never reduce the significant figures of a result, only measured quantities do.

Combining measurements. Two rules cover almost every situation in this book:

- **Multiplication and division:** the result has the same number of *significant figures* as the least precise input.
- **Addition and subtraction:** the result has the same number of *decimal places* as the least precise input.

For example, $(3.1\ \mathrm{m}) \times (3.8\ \mathrm{m}) = 11.78\ \mathrm{m}^2$, which we round to $12\ \mathrm{m}^2$ (two significant figures, matching the inputs). Carrying more digits than the data support gives a false impression of precision.

Example 1.1: Applying both rules to one rectangle

A rectangular plot is measured as $12.5\ \mathrm{m}$ long and $4.2\ \mathrm{m}$ wide. Report its area and its perimeter with the appropriate number of significant figures.

Solution

Count first. 12.5 has three significant figures and is known to the nearest tenth of a meter; 4.2 has two significant figures and is also known to the nearest tenth.

Area. The area is a product, so the multiplication rule applies: the result may keep only as many significant figures as the *less* precise input, which is two.

$$A = (12.5\,\mathrm{m})(4.2\,\mathrm{m}) = 52.5\,\mathrm{m}^2 \longrightarrow A = 53\,\mathrm{m}^2. \tag{1.3.1}$$

To keep two figures we keep the 5 and the 2 and look at the next digit, a 5; by the usual convention it rounds up, giving 53.

Perimeter. The perimeter is a sum, so the addition rule applies to the sum inside the parentheses:

$$P = 2(12.5\,\mathrm{m} + 4.2\,\mathrm{m}) = 2(16.7\,\mathrm{m}) = 33.4\,\mathrm{m}. \tag{1.3.2}$$

Both lengths are known to the nearest tenth of a meter, so their sum keeps one decimal place. The factor of 2 is exact (a rectangle has exactly two lengths and two widths), so it does not limit the result. The perimeter therefore keeps three significant figures.

Notice that the same two measurements support three significant figures in one result and only two in the other. The rules are not about the numbers themselves; they are about which operation you performed.

Note: Rounding, and a convention used throughout this book

Round only at the *end* of a calculation. Intermediate steps should carry every digit available to them, because rounding partway through a multi-step problem lets small errors accumulate and can shift the last digit of the final answer.

Your reported final answer, on the other hand, must carry exactly the number of significant figures the data support. That is the digit count a grader, a reader, or a collaborator will hold you to.

To keep the worked examples in this book readable, the values displayed in them are written to three significant figures for brevity, even where the arithmetic behind them carries more. When you report a final answer of your own, apply the rules above rather than copying the three-digit display.

Accuracy is not precision. *Precision* describes how finely a measurement is resolved and how repeatably an instrument returns the same value; *accuracy* describes how close that value lands to the truth. A scale that reads $2.4983\ \mathrm{kg}$ every single time for a $2.0000\ \mathrm{kg}$ standard is extremely precise and badly inaccurate: it needs calibration, not more digits. Significant figures communicate precision alone. They say nothing about whether the instrument was telling the truth.

Checkpoint 1.1

A digital scale reads $2.50\ \mathrm{kg}$. A second scale, with finer resolution, reads $2.4983\ \mathrm{kg}$. Which measurement has more significant figures? Are the two readings consistent?

Answer

Answers to Checkpoints

Checkpoint 1.1. The second reading, $2.4983\ \mathrm{kg}$, has five significant figures; the first, $2.50\ \mathrm{kg}$, has three. The two are consistent. The first scale resolves only to $0.01\ \mathrm{kg}$, so its reading means "somewhere between 2.495 and $2.505\ \mathrm{kg}$," and $2.4983\ \mathrm{kg}$ lies inside that range (it rounds to $2.50\ \mathrm{kg}$). The finer scale has not contradicted the coarser one; it has only narrowed down where in that range the mass lies.

This page titled 1.3: Scientific Notation and Significant Figures is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.4: Unit Conversions

Many calculations require converting a quantity from one unit to another: between SI prefixes (km to nm), between SI and non-SI units (meters to inches), or between compound units (Mm/s to mi/hr). The systematic technique is to multiply by one or more **conversion factors**: ratios equal to 1 that are constructed from known equivalences.

The conversion-factor method

If $1\ \text{m} = 100\ \text{cm}$, then both $\dfrac{100\,\text{cm}}{1\,\text{m}}$ and $\dfrac{1\,\text{m}}{100\,\text{cm}}$ equal 1. Multiplying any quantity by either of these does not change its value, only how it is expressed. The trick is to choose the orientation that makes the unwanted unit *cancel*, leaving the unit you want.

Example 1.2: A simple length conversion

How many inches are in $4.2\ \text{m}$? Use the equivalence $1\ \text{m} = 3.28\ \text{ft}$ and $1\ \text{ft} = 12\ \text{in}$.

Solution

Chain two conversion factors so that meters and feet both cancel:

$$4.2\ \cancel{\text{m}} \times \frac{3.28\ \cancel{\text{ft}}}{1\ \cancel{\text{m}}} \times \frac{12\,\text{in}}{1\ \cancel{\text{ft}}} = (4.2)(3.28)(12)\,\text{in}. \tag{1.4.1}$$

The meters cancel against the first ratio, the feet cancel against the second, and the only unit left is inches. Always check the cancellation *before* reaching for the calculator: if the wrong units survive, the conversion factors are upside down.

The arithmetic can be done one factor at a time: $4.2 \times 3.28 = 13.776$ (the length in feet), and $13.776 \times 12 = 165.3$ (the length in inches).

Now apply the significant-figure rules. The 12 is exact (a foot is *defined* as twelve inches), so it imposes no limit, and 3.28 has three figures. The measured length $4.2\ \text{m}$ has only two, so the answer keeps two:

$$165.3\,\text{in} \longrightarrow 1.7 \times 10^{2}\,\text{in}. \tag{1.4.2}$$

Writing "$170\ \text{in}$" would leave the reader unsure whether the zero is significant; scientific notation says two figures unambiguously.

Example 1.3: Football field in inches

How long, in inches, is a standard American football field, measured from goal line to goal line (100 yards)? (Including the two end zones the field is 120 yards long.)

Solution

Use $1\ \text{yd} = 3\ \text{ft}$ and $1\ \text{ft} = 12\ \text{in}$:

$$100\ \cancel{\text{yd}} \times \frac{3\ \cancel{\text{ft}}}{1\ \cancel{\text{yd}}} \times \frac{12\,\text{in}}{1\ \cancel{\text{ft}}} = 3600\,\text{in}. \tag{1.4.3}$$

Converting between SI prefixes

Conversions between prefixed SI units follow the same pattern; the conversion factor is simply a power of ten read off the prefix table.

Example 1.4: 140 km to nm

Convert $140\ \text{km}$ to nanometers.

Solution

Step 1: Build the conversion factors. The prefix table gives $1\ \text{km} = 10^{3}\ \text{m}$ directly. It gives nano- as 10^{-9}, which means $1\ \text{nm} = 10^{-9}\ \text{m}$. We want the factor the other way around (meters in terms of nanometers), so multiply both sides by 10^{9}:

$$10^{9} \times (1\,\text{nm}) = 10^{9} \times (10^{-9}\,\text{m}) \quad \Longrightarrow \quad 1\,\text{m} = 10^{9}\,\text{nm}, \tag{1.4.4}$$

since $10^{9} \times 10^{-9} = 10^{9+(-9)} = 10^{0} = 1$.

Step 2: Chain them. Orient each factor so the unwanted unit cancels:

$$140\ \cancel{\text{km}} \times \frac{10^{3}\ \cancel{\text{m}}}{1\ \cancel{\text{km}}} \times \frac{10^{9}\,\text{nm}}{1\ \cancel{\text{m}}} = 140 \times 10^{3} \times 10^{9}\,\text{nm}. \tag{1.4.5}$$

When powers of ten are multiplied, their exponents *add*: $10^{3} \times 10^{9} = 10^{3+9} = 10^{12}$.

Step 3: Put the result in scientific notation. 140×10^{12} is correct but not in standard form, because 140 is not between 1 and 10. Write $140 = 1.4 \times 10^{2}$ and add the exponents again:

$$140 \times 10^{12}\,\text{nm} = 1.4 \times 10^{2} \times 10^{12}\,\text{nm} = 1.4 \times 10^{14}\,\text{nm}. \tag{1.4.6}$$

Check. A nanometer is an extremely small unit, so it should take an enormous number of them to span $140\ \text{km}$. It does.

Example 1.5: 208 ps to Ts

Convert $208\ \text{ps}$ to teraseconds.

Solution

From the prefix table, $1\ \text{ps} = 10^{-12}\ \text{s}$ and $1\ \text{Ts} = 10^{12}\ \text{s}$. Because seconds must cancel, the second factor has seconds on the bottom:

$$208\ \cancel{\text{ps}} \times \frac{10^{-12}\ \cancel{\text{s}}}{1\ \cancel{\text{ps}}} \times \frac{1\,\text{Ts}}{10^{12}\ \cancel{\text{s}}} = 208 \times \frac{10^{-12}}{10^{12}}\,\text{Ts}. \tag{1.4.7}$$

When one power of ten is divided by another, the exponents *subtract*:

$$\frac{10^{-12}}{10^{12}} = 10^{-12-12} = 10^{-24}. \tag{1.4.8}$$

Watch the signs: it is tempting to see -12 and 12 and conclude that they cancel to 10^{0}, but that would be true only if we were *multiplying*. Finally, write $208 = 2.08 \times 10^{2}$ to reach standard form:

$$208 \times 10^{-24}\,\text{Ts} = 2.08 \times 10^{2} \times 10^{-24}\,\text{Ts} = 2.08 \times 10^{-22}\,\text{Ts}. \tag{1.4.9}$$

Check. A picosecond is a very short time, and a terasecond is a very long one: a year is about $3.16 \times 10^{7}\ \text{s}$, so $1\ \text{Ts} = 10^{12}\ \text{s}$ $\div\,(3.16 \times 10^{7}\ \text{s/yr}) \approx 31\ 700$ years. It makes sense that the numerical answer is tiny.

Converting units raised to powers

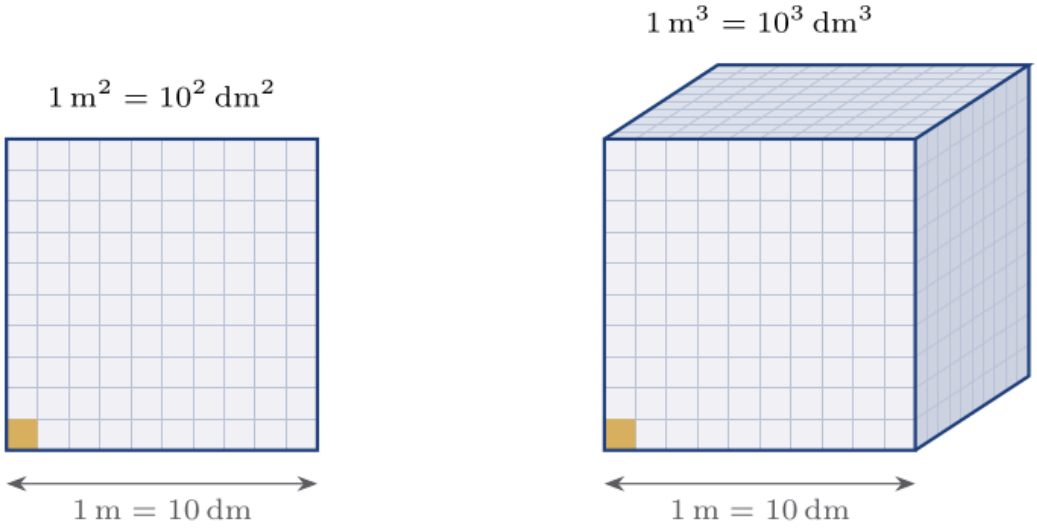


Figure 1.2. Why a conversion factor must be raised to the same power as its unit. Dividing each edge into ten parts produces 10^2 small squares and 10^3 small cubes, not ten of either; one of each is shaded. The same argument with 100 divisions per edge (centimeters instead of decimeters) gives $1\ \mathrm{m}^2 = 10^4\ \mathrm{cm}^2$ and $1\ \mathrm{m}^3 = 10^6\ \mathrm{cm}^3$. As a bonus, one of the small cubes on the right is a cubic decimeter, better known as a liter.

•Figure description

Two panels comparing area and volume scaling. The left panel shows a square with each edge divided into ten equal parts, forming a ten-by-ten grid of one hundred small squares, with a single small square in the corner shaded gold. The right panel shows a cube drawn in perspective with each edge likewise divided into ten, containing one thousand small cubes, with one small cube shaded. Dividing every edge into ten parts yields ten squared small squares and ten cubed small cubes, not ten of either.

When the unit being converted is squared, cubed, or otherwise raised to a power, the conversion factor must be raised to the same power. This is the single most common source of error in this chapter, and it is worth seeing geometrically rather than algebraically. Figure 1.2 shows why: a square meter is not 100 square centimeters but 100×100 of them, and a cubic meter is not 100 cubic centimeters but $100 \times 100 \times 100$.

Example 1.6: Area conversion

Convert $14\ \mathrm{m}^2$ to square centimeters.

Solution

Step 1: Square the length equivalence. Start from $1\ \mathrm{m} = 100\ \mathrm{cm}$ and square both sides. The exponent applies to everything inside the parentheses, the number *and* the unit, just as $(ab)^2 = a^2b^2$:

$$(1\,\mathrm{m})^2 = (100\,\mathrm{cm})^2 \quad \Longrightarrow \quad 1\,\mathrm{m}^2 = 100^2\,\mathrm{cm}^2 = 10\,000\,\mathrm{cm}^2. \tag{1.4.10}$$

In powers of ten, $100^2 = (10^2)^2 = 10^4$.

Step 2: Convert.

$$14\ \cancel{\mathrm{m}^2} \times \frac{10^4\,\mathrm{cm}^2}{1\ \cancel{\mathrm{m}^2}} = 14 \times 10^4\,\mathrm{cm}^2 = 1.4 \times 10^5\,\mathrm{cm}^2, \tag{1.4.11}$$

where the last step writes $14 = 1.4 \times 10^1$ and adds exponents.

The common mistake. Using the unsquared factor gives $14 \times 100 = 1400\ \mathrm{cm}^2$, which is too small by a factor of 100. The quick test: did the unit m^2 actually cancel? With the factor $100\ \mathrm{cm}/1\ \mathrm{m}$ it would not; you would be left with $\mathrm{m} \cdot \mathrm{cm}$.

Example 1.7: Volume conversion

Convert $4.19 \times 10^{24}\ \text{mm}^3$ to cubic megameters.

Solution

Step 1: Cube the length equivalences. From the prefix table, $1\ \text{mm} = 10^{-3}\ \text{m}$ and $1\ \text{Mm} = 10^{6}\ \text{m}$. Cube both sides of each, remembering that the cube applies to the number and the unit. A power of a power *multiplies* the exponents, $(10^{a})^{b} = 10^{ab}$:

$$\begin{aligned} 1\,\text{mm}^3 &= (10^{-3}\,\text{m})^3 = 10^{(-3)(3)}\,\text{m}^3 = 10^{-9}\,\text{m}^3, \\ 1\,\text{Mm}^3 &= (10^{6}\,\text{m})^3 = 10^{(6)(3)}\,\text{m}^3 = 10^{18}\,\text{m}^3. \end{aligned} \tag{1.4.12}$$

Step 2: Chain the factors. Go from mm^3 to m^3, then from m^3 to Mm^3:

$$4.19 \times 10^{24}\ \cancel{\text{mm}^3} \times \frac{10^{-9}\ \cancel{\text{m}^3}}{1\ \cancel{\text{mm}^3}} \times \frac{1\,\text{Mm}^3}{10^{18}\ \cancel{\text{m}^3}} = 4.19 \times \frac{10^{24} \times 10^{-9}}{10^{18}}\,\text{Mm}^3. \tag{1.4.13}$$

Step 3: Collect the powers of ten. Exponents in the numerator add; the exponent in the denominator subtracts:

$$\frac{10^{24} \times 10^{-9}}{10^{18}} = 10^{24+(-9)-18} = 10^{-3}, \qquad \text{so} \qquad 4.19 \times 10^{24}\,\text{mm}^3 = 4.19 \times 10^{-3}\,\text{Mm}^3. \tag{1.4.14}$$

Check. The number went from 10^{24} to 10^{-3}, a drop of 27 powers of ten. That is exactly the ratio of the two volume units: a megameter is 10^{9} millimeters, so a cubic megameter is $(10^{9})^{3} = 10^{27}$ cubic millimeters.

Converting compound units

When a quantity has multiple units (a speed in meters per second, a density in grams per cubic centimeter), convert each unit separately. It's often clearest to lay every conversion factor out as a chain.

Example 1.8: Speed conversion

Convert $45\ \text{Mm/s}$ to miles per hour, using $1\ \text{mi} = 1609\ \text{m}$ and $1\ \text{hr} = 3600\ \text{s}$.

Solution

Step 1: Convert the numerator and the denominator separately. The length unit is on top, so its factors must put Mm on the bottom and then m on the bottom. The time unit is on the *bottom* of Mm/s, so to cancel it the time factor must put seconds on the *top*: $3600\ \text{s}/1\ \text{hr}$, not $1\ \text{hr}/3600\ \text{s}$.

$$45\ \frac{\cancel{\text{Mm}}}{\cancel{\text{s}}} \times \frac{10^{6}\ \cancel{\text{m}}}{1\ \cancel{\text{Mm}}} \times \frac{1\,\text{mi}}{1609\ \cancel{\text{m}}} \times \frac{3600\ \cancel{\text{s}}}{1\,\text{hr}} = \frac{45 \times 10^{6} \times 3600}{1609}\ \frac{\text{mi}}{\text{hr}}. \tag{1.4.15}$$

It makes physical sense that converting "per second" to "per hour" *multiplies* by 3600: in an hour the object covers 3600 times the distance it covers in a second.

Step 2: Do the arithmetic. The numerator is $45 \times 3600 = 162\ 000 = 1.62 \times 10^{5}$, so $45 \times 10^{6} \times 3600 = 1.62 \times 10^{5} \times 10^{6} = 1.62 \times 10^{11}$. Dividing,

$$\frac{1.62 \times 10^{11}}{1609} = 1.007 \times 10^{8}\,\text{mi/hr} \approx 1.0 \times 10^{8}\,\text{mi/hr}, \tag{1.4.16}$$

rounded to two significant figures to match the 45.

Check. Light travels $c = 3.00 \times 10^{8}\ \text{m/s} = 300\ \text{Mm/s}$, which the same chain of conversions turns into about $6.7 \times 10^{8}\ \text{mi/hr}$. Our object is moving at $45/300 = 0.15$ of that, and $0.15 \times 6.7 \times 10^{8} \approx 1.0 \times 10^{8}\ \text{mi/hr}$. The answer survives the check.

Example 1.9: Filling a swimming pool

How long will it take to fill an Olympic swimming pool of dimensions $50\ \text{m} \times 25\ \text{m} \times 2\ \text{m}$ from a garden hose delivering 15 gallons per minute? Use $1\ \text{m}^3 = 264.172\ \text{gal}$.

Solution

Step 1: Find the volume. For a rectangular box, volume is length times width times depth:

$$V = (50\,\text{m})(25\,\text{m})(2\,\text{m}) = 2500\,\text{m}^3. \tag{1.4.17}$$

Step 2: Put the volume in the hose's units. The flow rate is in gallons, so convert:

$$V = 2500\,\cancel{\text{m}^3} \times \frac{264.172\,\text{gal}}{1\,\cancel{\text{m}^3}} = 660\,430\,\text{gal} \approx 6.60 \times 10^5\,\text{gal}. \tag{1.4.18}$$

Step 3: Solve for the time. A flow rate is volume delivered per unit time, $R = V/t$. We want t, so multiply both sides by t and then divide both sides by R:

$$R = \frac{V}{t} \quad \Longrightarrow \quad Rt = V \quad \Longrightarrow \quad t = \frac{V}{R}. \tag{1.4.19}$$

Substituting,

$$t = \frac{6.60 \times 10^5\,\text{gal}}{15\,\text{gal/min}} \approx 4.40 \times 10^4\,\text{min}. \tag{1.4.20}$$

The units work out because dividing by a fraction is the same as multiplying by its reciprocal: $\text{gal} \div (\text{gal/min}) = \text{gal} \times (\text{min/gal})$ $= \text{min}$. If you had divided the other way, the units would have come out as $1/\text{min}$: a sure sign of an error.

Step 4: Express the answer in sensible units. Tens of thousands of minutes is hard to picture, so convert to days:

$$4.40 \times 10^4\,\cancel{\text{min}} \times \frac{1\,\cancel{\text{hr}}}{60\,\cancel{\text{min}}} \times \frac{1\,\text{day}}{24\,\cancel{\text{hr}}} \approx 31\,\text{days}. \tag{1.4.21}$$

A garden hose running continuously for a month would just barely do it, which is why pools are filled from water mains, not hoses.

This page titled 1.4: Unit Conversions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.5: Dimensional Analysis and Estimation

Checking equations with dimensions

Even before plugging in numbers, you can often catch errors in an equation by checking its **dimensions**: the kind of physical quantity each term represents, independent of which unit it's expressed in. Length, mass, and time are the three fundamental dimensions in mechanics; we denote them $[\mathrm{L}]$, $[\mathrm{M}]$, and $[\mathrm{T}]$. Every term in a physical equation must have the same combination of dimensions, and the argument of any transcendental function (sine, exponential, logarithm) must be dimensionless.

For instance, the kinematic equation $\Delta x = v_0\ t + \frac{1}{2} a\ t^2$ from Chapter 2 has

$$[\Delta x] = [\mathrm{L}], \qquad [v_0\, t] = \frac{[\mathrm{L}]}{[\mathrm{T}]} \cdot [\mathrm{T}] = [\mathrm{L}], \qquad [a\, t^2] = \frac{[\mathrm{L}]}{[\mathrm{T}]^2} \cdot [\mathrm{T}]^2 = [\mathrm{L}]. \tag{1.5.1}$$

Every term has dimension $[\mathrm{L}]$, so the equation is dimensionally consistent. If you ever derive an equation in which the terms don't match dimensionally, the equation is wrong: no further checking required.

The converse, however, does not hold, and this is worth being clear about. Dimensional consistency is a test an equation must *pass*, not a certificate that it is correct. Suppose you misremembered the kinematic equation as $\Delta x = v_0 t + 2at^2$: every term still has dimension $[\mathrm{L}]$, so the check reports nothing wrong, yet the equation is false. Dimensional analysis catches errors in the s*tructure* of an equation (a missing factor of t, an acceleration where a velocity belongs) and is blind to errors in the pure numbers. It is a cheap test that rules results out; it never rules them in.

Checkpoint 1.2

A student writes the equation $v^2 = v_0 + 2\ a\ \Delta x$ for one-dimensional motion with constant acceleration. Is this equation dimensionally consistent? If not, identify the term that is incorrect.

Answer

Order-of-magnitude estimation

Sometimes the useful question is not "what is the answer to three digits?" but "is this closer to a hundred or to a million?" An **order-of-magnitude estimate** answers a question to within roughly a factor of ten, using rough inputs and deliberately crude arithmetic. Working physicists and engineers do this constantly, to decide whether an effect is large enough to bother measuring, to size a piece of equipment before designing it, and to check that a long calculation has not gone off by a factor of a thousand.

The method is simple. Break the quantity into factors you can each guess to within a factor of a few, round each factor to one significant figure or to the nearest power of ten, multiply, and report the result as a single power of ten. Individual guesses that are too high tend to be offset by others that are too low, so the estimate is usually far better than its ingredients deserve.

Example 1.10: Order-of-magnitude estimate: the air in a room

Roughly what is the mass of the air in a typical classroom?

Solution

Step 1: Estimate the volume. Estimate the room as about $10\ \mathrm{m}$ by $8\ \mathrm{m}$, with a $3\ \mathrm{m}$ ceiling:

$$V \approx (10\,\mathrm{m})(8\,\mathrm{m})(3\,\mathrm{m}) = 240\,\mathrm{m}^3. \tag{1.5.2}$$

Step 2: Relate mass to volume. Density is mass per unit volume, $\rho = m/V$. Multiplying both sides by V gives the mass, $m = \rho V$. The density of air at ordinary room conditions is about $1.2\ \mathrm{kg/m^3}$: roughly a thousandth the density of water. Then

$$m \approx (1.2\,\mathrm{kg/m^3})(240\,\mathrm{m^3}) \approx 290\,\mathrm{kg} \approx 3 \times 10^2\,\mathrm{kg}. \tag{1.5.3}$$

The cubic meters cancel, leaving kilograms. Round only this last result to a single figure; the inputs were rough, so one figure is all the answer can claim.

The air in the room is invisible and feels like nothing at all, yet it amounts to some $300\ \mathrm{kg}$: about the combined mass of four adults.

Note carefully what this estimate does and does not claim. It does not claim $288\ \mathrm{kg}$; the room was never measured. It claims that the answer is a few hundred kilograms rather than three kilograms or three tonnes, and it reaches that conclusion in two lines. That is frequently all you need, and it is always worth having before you commit to a longer calculation.

Answers to Checkpoints

Checkpoint 1.2. No. The left side, v^2, has dimensions $[L^2/T^2]$, and so does the last term: $2a\ \Delta x$ is $[L/T^2] \times [L] = [L^2/T^2]$ (the 2 is a pure number). The middle term v_0, however, has dimensions $[L/T]$, which cannot be added to $[L^2/T^2]$. The incorrect term is v_0; it should be v_0^2, giving the correct equation $v^2 = v_0^2 + 2a\ \Delta x$.

This page titled 1.5: Dimensional Analysis and Estimation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.S: Measurement (Summary)

Summary

- **Physics** studies matter, energy, space, and time. A physics problem differs from a math problem by requiring you to model a real situation before applying formulas.
- Every measurement is a comparison with a **standard**. The SI base units used in mechanics are the **meter** (m), the **kilogram** (kg), and the **second** (s); other units, like the **newton** ($\mathrm{kg \cdot m/s^2}$), are derived from these.
- **Scientific notation**, $N = a \times 10^n$ with $1 \leq |a| < 10$, lets us write very large and very small numbers compactly. **SI prefixes** (kilo-, milli-, micro-, etc.) provide an equivalent shorthand attached directly to the unit.
- **Significant figures** represent the precision of a measured or computed value. Multiplication/division: match the least precise input's *sig figs*. Addition/subtraction: match the least precise input's *decimal places*. Exact numbers and defined conversion factors impose no limit. Round only the final answer; carry full precision through intermediate steps.
- **Precision** (how finely resolved and how repeatable) is not **accuracy** (how close to the truth). Significant figures report the first and are silent about the second.
- **Unit conversions** are performed by multiplying by ratios equal to 1, oriented so that the unwanted units cancel. For powers (areas, volumes), raise the entire conversion factor to the appropriate power.
- **Dimensional analysis** requires every term in an equation to share the same combination of base dimensions $[\mathrm{L}]$, $[\mathrm{M}]$, $[\mathrm{T}]$. It is a powerful, cheap check that can rule a result *out*, but a dimensionally consistent equation may still be wrong.
- An **order-of-magnitude estimate** answers a question to within about a factor of ten from rough inputs. It tells you the size of an answer before you commit to computing it exactly.

Key Equations

Definition of the newton

$$1\,\mathrm{N} = 1\,\frac{\mathrm{kg \cdot m}}{\mathrm{s^2}} \tag{1.S.1}$$

Scientific notation

$$N = a \times 10^n, \qquad 1 \leq |a| < 10, \quad n \in \mathbb{Z} \tag{1.S.2}$$

Significant figures

- **Multiplication/division:** match the least precise input's number of *significant figures*.
- **Addition/subtraction:** match the least precise input's number of *decimal places*.
- Exact numbers and defined conversion factors never limit significant figures.

Dimensional analysis

- Every term in a valid equation must share the same combination of dimensions $[\mathrm{L}]$, $[\mathrm{M}]$, $[\mathrm{T}]$.
- Passing this check rules nothing *in*: it can only rule an equation *out*.

This page titled 1.S: Measurement (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

1.E: Measurement (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 1.1 — What is Physics?

1.1 ★ *[Conceptual.]* For each of the following questions, state whether you think it is best answered by mathematics, physics, or a different field of science, and briefly justify your choice.

a. How long does it take a dropped ball to fall $1.5\ \text{m}$?
b. What is the largest prime number less than 1000?
c. Why does ice float on liquid water?
d. Why do humans dream?

1.2 ★★ *[Conceptual.]* Engineering students sometimes describe physics as "just math with units." Write a one-paragraph response either supporting or rebutting this characterization, with at least one specific example.

Section 1.2 — Measuring Standards

1.3 ★ Express each of the following in terms of the SI base units (m, kg, s):

a. the joule, $1\ \text{J} = 1\ \text{N}\cdot\text{m}$;
b. the watt, $1\ \text{W} = 1\ \text{J/s}$;
c. the pascal, $1\ \text{Pa} = 1\ \text{N/m}^2$.

1.4 ★★ Since 1983 the meter has been defined by fixing the speed of light in vacuum at exactly $299\ 792\ 458\ \text{m/s}$, with no uncertainty whatsoever.

a. Using this definition, how many meters does light travel in $1.0\ \text{ns}$?
b. Approximately how long does it take light to travel from the Sun to the Earth, given the Earth–Sun distance of about $1.50 \times 10^{11}\ \text{m}$ (one astronomical unit, $1\ \text{au}$)?
c. In one beat of a human heart ($\sim 0.85\ \text{s}$), how far does light travel? Compare this distance to the Earth–Sun distance from part (b).

1.5 ★ *[Conceptual.]* In 2019, the International Prototype of the Kilogram (a physical platinum–iridium cylinder kept in a vault near Paris) was retired in favor of a definition of the kilogram based on the Planck constant. Give at least two reasons why defining a unit in terms of a fundamental physical constant is preferable to defining it in terms of a physical artifact.

1.6 ★★ When the meter was first defined in 1791, it was set equal to one ten-millionth of the distance from the North Pole to the equator, measured along a meridian.

a. Treating the Earth as a perfect sphere, use this definition to find the Earth's circumference along a meridian, in kilometers.
b. From your answer to part (a), find the Earth's radius. Compare it with the accepted mean radius of $6.371 \times 10^{6}\ \text{m}$ by computing the percent difference.
c. Modern measurements give the pole-to-equator distance as about $1.0002 \times 10^{7}\ \text{m}$. If the 1791 definition had been carried out perfectly, how long would the meter be, in today's meters? By how many millimeters does this differ from the meter we actually use?
d. *[Conceptual.]* When the meter was redefined in 1960 and again in 1983, the new definitions were chosen to match the existing meter rather than to correct it back to one ten-millionth of the true pole-to-equator distance. Why?

Section 1.3 — Scientific Notation and Significant Figures

1.7 ★ Rewrite each quantity using the most appropriate SI prefix (so the leading coefficient is between 1 and 1000):

a. $4.5 \times 10^{-9}\ \text{m}$
b. $3.2 \times 10^{6}\ \text{s}$
c. $7.1 \times 10^{-3}\ \text{A}$
d. $8.4 \times 10^{12}\ \text{Hz}$

1.8 ★★ State the number of significant figures in each of the following measurements, and rewrite each in scientific notation:

a. $0.0027\ \text{m}$
b. $4.500\ \text{kg}$
c. $1200\ \text{s}$ *(ambiguous: comment on what assumption you must make)*
d. 6.022×10^{23} (the Avogadro number, quoted here to four sig figs)

1.9 ★★ Carry out the following calculations and report the result with the correct number of significant figures.

a. $(2.34\ \text{m}) \times (1.7\ \text{m})$
b. $\dfrac{12.45\,\text{kg}}{3.2\,\text{s}}$
c. $(15.6\ \text{m}) + (2.04\ \text{m}) - (0.083\ \text{m})$
d. $\sqrt{(3.0\,\text{m})^2 + (4.0\,\text{m})^2}$

1.10 ★ *[Conceptual.]* A student measures a length and reports the result as $0.5\ \text{m}$. A second student, working independently, reports the same length as $0.500\ \text{m}$. Are these two reports equivalent? If not, what is the practical difference between them? Which student's instrument was more precise?

Section 1.4 — Unit Conversions

1.11 ★ Carry out the following conversions:

a. $75\ \text{mi/hr}$ to m/s.
b. $60\ \text{mi/hr}$ to km/hr.
c. $1.0\ \text{L}$ to cm^3. *(Hint:* $1\ \text{L} = 1000\ \text{cm}^3$*.)*
d. $9.81\ \text{m/s}^2$ to ft/s^2, using $1\ \text{m} = 3.281\ \text{ft}$.

1.12 ★ The density of water is approximately $1000\ \text{kg/m}^3$. Convert this to g/cm^3. Does your answer match a value you may already know?

Section 1.5 — Dimensional Analysis and Estimation

1.13 ★ Check each of the following equations for dimensional consistency. State whether each is dimensionally correct, and if not, identify the offending term. (Symbols: v and v_0 are velocities; a is acceleration; t is time; x is position; m is mass; F is force; E is energy.)

a. $v = v_0 + a\ t$
b. $x = v_0\ t + \frac{1}{2} a\ t^2$
c. $v^2 = v_0^2 + 2\ a$
d. $m\ v = m\ v_0 + a\ t$
e. $E = \frac{1}{2} m\ v^2$
f. $F = m\ a + m\ v$

1.14 ★ *[Conceptual.]* A classmate claims that dimensional analysis can be used to *prove* that an equation is correct. Explain why this is not quite true: give an example of an equation that is dimensionally consistent but still physically wrong. What *can* dimensional analysis reliably tell you?

Challenge Problems

1.15 ★★ A laser pulse has a duration of $15\ \text{fs}$ (femtoseconds). Light travels at $c = 3.00 \times 10^8\ \text{m/s}$ in vacuum.

a. How far does light travel during the pulse? Express your answer in micrometers.
b. How does that distance compare to the diameter of a human hair ($\sim 75\ \mu\text{m}$)?
c. If the pulse repetition rate is $80\ \text{MHz}$, how many pulses are produced each second? Hint: $1\ \text{Hz} = 1\ \text{s}^{-1}$

1.16 ★★ *[Real-world application: sports.]* Usain Bolt's world-record $100\ \text{m}$ sprint took $9.58\ \text{s}$.

a. What was his average speed in m/s?
b. Convert that speed to mi/hr.
c. Bolt is reported to have hit a peak speed of about $12.4\ \text{m/s}$ mid-race. Express this peak speed in km/hr.
d. At Bolt's peak speed, how long would it take him to traverse a $400\ \text{m}$ track (assuming he could hold it)? Compare to the actual world record of $43.03\ \text{s}$.

1.17 ★★ *[Real-world application: engineering.]* A residential solar panel produces about $350\ \text{W}$ of electrical power under full sunlight.

a. Express this in kilowatts.
b. Over a $6.0\ \text{hr}$ day of full sunlight, how many kilowatt-hours of energy does the panel produce? Note: $1\ \text{kW}\cdot\text{hr} = 3.6\times10^{6}\ \text{J}$.
c. Express the daily energy production in joules using scientific notation.
d. A typical American household uses about $30\ \text{kW}\cdot\text{hr}$ of electricity per day. How many of these solar panels (operating $6.0\ \text{hr/day}$) would be needed to meet that demand?

1.18 ★★ *[Real-world application: biology.]* A typical adult's heart beats roughly 70 times per minute, pumping about $70\ \text{mL}$ of blood per beat.

a. About how many heartbeats occur in an 80-year lifetime? Express in scientific notation.
b. About how many liters of blood does the heart pump in a lifetime?
c. Compare that volume to the volume of an Olympic swimming pool ($2500\ \text{m}^3$). Express the ratio as a number to one significant figure.

1.19 ★★★ *[The dorm-room wallpaper problem.]* A student's dormitory room measures $3.1\ \text{m}\times3.8\ \text{m}$, with a $2.6\ \text{m}$ ceiling. The student wallpapers all four walls (including the door and window) with the 480 sheets of their physics textbook, each $21\ \text{cm}\times28\ \text{cm}$.

a. Compute the total wall area in m^2.
b. Compute the total area of the 480 textbook sheets in m^2.
c. Show that the situation described is impossible by comparing your two answers.
d. By what factor would the total *area* of the sheets have to increase for the scenario to work? If instead the student shrank the room, keeping its shape, by what factor would each *length* have to shrink?

1.20 ★★★ *[Real-world application: engineering.]* A water tank in the shape of a vertical cylinder has a radius of $1.5\ \text{m}$ and is filled to a height of $4.0\ \text{m}$. The density of water is $1000\ \text{kg/m}^3$. Note: the volume of a cylinder is $V=\pi r^2 h$, and $1\ \text{m}^3 = 264.172\ \text{gal}$.

a. Compute the volume of water in m^3, then convert it to gallons.
b. What is the mass of the water in kilograms? Express in scientific notation.
c. If the tank drains at a steady rate of $3.2\ \text{gal/min}$, how long (in hours) will it take to empty?
d. Convert the drain rate to m^3/s.

1.21 ★★★ A sheet of paper has a thickness of about $0.10\ \text{mm}$. If you could fold a single sheet in half n times (a feat physically impossible past about $n=8$ for an ordinary sheet, though in 2002 a high-school student named Britney Gallivan managed 12 folds using a single strip of tissue paper nearly $1.2\ \text{km}$ long), the resulting stack has a thickness $t_n = t_0\cdot 2^n$, where t_0 is the original thickness.

a. After 20 folds, how thick is the stack? Express in meters.
b. After 42 folds, the stack would reach the Moon. Verify this, using an Earth–Moon distance of $3.84\times10^{8}\ \text{m}$.
c. After how many additional folds (beyond 42) would the stack reach the Sun, at $1.50\times10^{11}\ \text{m}$?

1.22 ★★★ A simple pendulum consists of a small bob of mass m swinging on a string of length L near Earth's surface, where the free-fall acceleration is g. We suspect its period T (the time for one full swing) depends only on L, m, and g, in the form $T = k\,L^{a}m^{b}g^{c}$, where k is a dimensionless constant.

a. By demanding that both sides have the same dimensions, write equations for the exponents a, b, and c, and solve them.
b. Use your result to express T. Which of the three quantities does the period *not* depend on?
c. Dimensional analysis fixes the exponents but leaves k completely undetermined. Look up or recall the exact period of a simple pendulum: what is the actual value of k?
d. Using your formula, by what factor does the period change if the length is quadrupled? What if the same pendulum is taken to the Moon, where g is about one-sixth of its Earth value?

1.23 ★★★ *[Real-world application: engineering.]* In the United States fuel economy is quoted in miles per gallon (mpg); a certain car is rated at $30\ \mathrm{mi/gal}$. Surprisingly, the *fuel consumption* (the reciprocal, volume of fuel per distance) has the dimensions of an area. Use $1\ \mathrm{gal} = 3.785\ \mathrm{L}$ and $1\ \mathrm{mi} = 1609\ \mathrm{m}$.

a. Show that the dimensions of (gallons per mile) reduce to a length squared.
b. Compute the fuel consumption of this car as an area, in m^2.
c. European fuel consumption is quoted in liters per $100\ \mathrm{km}$. Convert $30\ \mathrm{mi/gal}$ to $\mathrm{L/100km}$, and confirm it corresponds to the same area as in (b).
d. Interpret the area physically: if all the fuel the car burns were squeezed into a thin thread laid along the road, what would the diameter of that thread be?

1.24 ★★★ *[Real-world application: estimation.]* A classic way to estimate the size of a molecule (used by Rayleigh and, earlier, Franklin) is to let a tiny drop of oil spread into a film on water. Suppose a drop of olive oil of volume $6.0 \times 10^{-5}\ \mathrm{cm}^3$ spreads into a roughly circular patch $20\ \mathrm{cm}$ in diameter, and assume the film is exactly one molecule thick.

a. Convert the drop's volume to m^3 and find the area of the film in m^2.
b. Estimate the thickness of the film, and hence the approximate length of one oil molecule. Express your answer in nanometers.
c. A single atom is about $0.1\ \mathrm{nm}$ across. Roughly how many atoms long is the oil molecule? Is this reasonable for a long-chain organic molecule?
d. State two assumptions built into this estimate that limit its accuracy.

This page titled 1.E: Measurement (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2: MOTION IN ONE DIMENSION

How do we describe the motion of a sprinter, a falling pinecone, or a jet slowing to a stop on an aircraft carrier? The branch of physics that answers such questions (without yet asking *why* an object moves the way it does) is called **kinematics**. In this chapter we restrict ourselves to motion along a single straight line, which lets us develop the essential vocabulary (position, displacement, velocity, acceleration), learn to read that vocabulary off a graph, and derive the four kinematic equations that govern motion under constant acceleration. We finish by applying the same machinery to one of the most important special cases in introductory physics: free fall under gravity near Earth's surface.

Learning Objectives

- Distinguish between distance and displacement, and between speed and velocity.
- Compute average and instantaneous velocities and accelerations from data, from graphs, and from the derivative definitions.
- Read velocity off the slope of a position–time graph, acceleration off the slope of a velocity–time graph, and displacement off the signed area under a velocity–time graph.
- Use the four kinematic equations to solve problems involving constant acceleration, and recognize when they do not apply.
- Apply these tools to objects in free fall, with the sign of g handled correctly.

Note: Notation in this chapter

Position, velocity, and acceleration are all *vector* quantities. In one dimension, though, a vector has only two possible directions, so we do not need boldface arrows to keep track of them: an ordinary signed number will do. A positive value means the $+x$ direction; a negative value means the $-x$ direction. Throughout this chapter, then, x, v, and a are *signed scalars*, and the sign *is* the direction. Full vector notation returns in Chapter 3, where motion is free to point anywhere in a plane.

2.1: Position, Time, and Displacement

2.2: Speed and Velocity

2.3: Acceleration

2.4: Motion Graphs

2.5: Kinematics in One Dimension

2.6: Free Fall

2.S: Motion in One Dimension (Summary)

2.E: Motion in One Dimension (Problems)

This page titled 2: Motion in One Dimension is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.1: Position, Time, and Displacement

To describe motion we first need a way to describe *location*. The **position** of a particle is its coordinate along the chosen axis. Along a one-dimensional axis (say, the x-axis) the position is a single number x that may be positive, negative, or zero, depending on which side of the origin the particle lies.

Before any of this means anything, you must make two choices: where to put the **origin** ($x = 0$) and which way to call **positive**. Both choices are entirely up to you, and no physical result depends on them, but every number you write down afterward does. Two students who choose differently will get answers that differ in sign while describing exactly the same motion. State your choice at the start of a problem and hold to it.

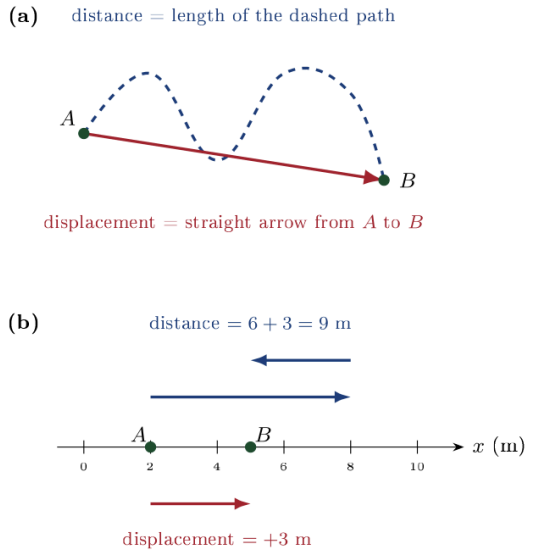


Figure 2.1. Distance versus displacement. **(a)** In two dimensions, distance is the length of the curved path actually traveled from A to B, while displacement is the straight arrow from A to B. **(b)** The version that matters for this chapter: a particle starts at A ($x = 2$ m), travels out to $x = 8$ m, then reverses and comes back to B ($x = 5$ m). It has traveled 9 m of distance but has been displaced only 3 m. In one dimension the two quantities differ precisely when the particle turns around.

•Figure description

Two panels contrasting distance with displacement. In the left panel a blue dashed curve wanders up and down from point A to point B, representing the actual path traveled, while a straight red arrow runs directly from A to B representing the displacement. In the right panel the same comparison is made along a single straight line, where the path doubles back on itself so the distance traveled exceeds the net displacement.

Checkpoint 2.1

Can a position be negative? Why is the answer different from the same question asked about distance?

Answer

When a particle moves from an initial position x_A to a final position x_B, its **displacement** is the change in position,

Key Equation: Displacement (1D)

$$\Delta x = x_B - x_A. \tag{2.1}$$

Displacement is a *vector* quantity: in one dimension it carries a sign that indicates direction along the axis. The **distance** traveled, by contrast, is the total length of the path the particle followed and is never negative (Figure 2.1). Distance ignores direction entirely; it simply accumulates.

The two agree only when the particle never reverses direction. The moment it turns around, the path keeps adding up while the displacement starts shrinking, which is why a runner who completes one lap of a track has covered $400\ \text{m}$ of distance and exactly zero displacement.

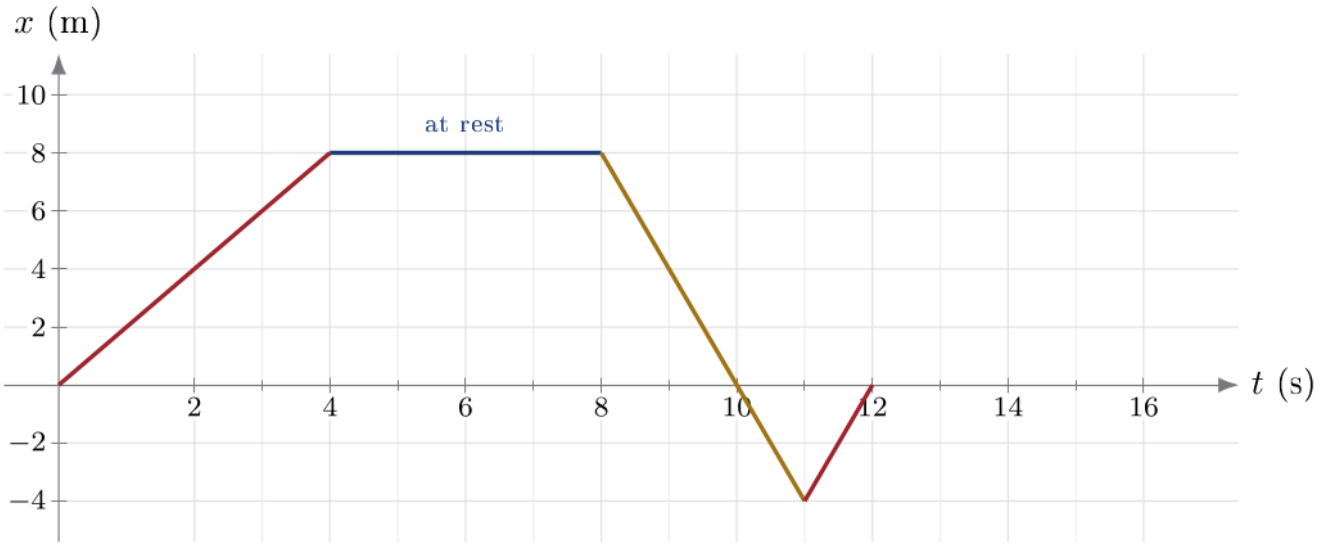


Figure 2.2. Position x of a particle as a function of time t. Red segments have positive slope (motion in the $+x$ direction), the blue segment has zero slope (the particle is at rest), and the gold segment has negative slope (motion in the $-x$ direction). Real objects cannot change velocity instantaneously, so the sharp corners here are an idealization; treat them as very rapid changes compressed to a point.

•Figure description

A position-versus-time graph, with time in seconds on the horizontal axis and position in meters on the vertical axis. The curve is built from straight segments of differing slope: red segments rise with positive slope, indicating motion in the positive x direction; a blue segment is horizontal, indicating the particle is momentarily at rest; and a gold segment falls with negative slope, indicating motion in the negative x direction. The slope at any point is the velocity.

Checkpoint 2.2

Using Figure 2.2, find

a. the displacement of the particle between $t = 0\ \text{s}$ and $t = 12\ \text{s}$;
b. the distance traveled by the particle over the same interval.

Answer

Answers to Checkpoints

Checkpoint 2.1. Yes. A position is a coordinate measured from an origin you chose, and a negative value simply means the object is on the $-x$ side of that origin; move the origin and the sign can change. Distance is different: it is the total length of path covered, which accumulates whichever way the object moves and never depends on the origin, so it can never be negative.

Checkpoint 2.2. (a) The particle starts at $x = 0$ at $t = 0$ and is back at $x = 0$ at $t = 12\ \text{s}$, so the displacement is $\Delta x = 0 - 0 = 0$. (b) The distance adds up every leg regardless of direction: out from 0 to $8\ \text{m}$ ($8\ \text{m}$), at rest, back from $8\ \text{m}$ to $-4\ \text{m}$ ($12\ \text{m}$), then forward from $-4\ \text{m}$ to 0 ($4\ \text{m}$). The total is $8 + 12 + 4 = 24\ \text{m}$: a large distance for zero displacement.

This page titled 2.1: Position, Time, and Displacement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.2: Speed and Velocity

The rate at which a particle's position changes is its **velocity**. Over a finite time interval Δt, the *average velocity* is the ratio of the displacement to the elapsed time,

Key Equation: Average velocity

$$v_{\text{avg}} = \frac{\Delta x}{\Delta t} = \frac{x_B - x_A}{t_B - t_A}. \tag{2.2}$$

Like displacement, velocity carries a sign in one dimension: positive when motion is in the $+x$ direction, negative when it is in the $-x$ direction.

The **average speed** is defined similarly, but in terms of total distance rather than displacement,

$$s_{\text{avg}} = \frac{\text{total distance traveled}}{\Delta t}. \tag{2.3}$$

Because distance is never negative, neither is average speed. Velocity is a vector and speed is a scalar, and they differ in exactly the way displacement and distance differ: velocity asks *where did you end up*, speed asks *how much ground did you cover*. The lap runner of the previous section had an average velocity of zero and an average speed that was decidedly not zero.

Instantaneous Velocity

The average velocity tells us the overall rate of motion across an interval but says nothing about what happens *at a particular instant*. To capture the latter we let $\Delta t \to 0$ in equation (2.2) and obtain the **instantaneous velocity**,

Key Equation: Instantaneous velocity

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{\mathrm{d}x}{\mathrm{d}t}. \tag{2.4}$$

Read $\frac{\mathrm{d}x}{\mathrm{d}t}$ as "the rate at which x is changing per unit of time, measured over an interval so short that the rate has no room to change within it." It is the average velocity of equation (2.2), computed over a vanishingly small window. Geometrically, the instantaneous velocity is the *slope of the tangent line* to the position–time graph at a single point: zoom in far enough on any smooth curve and it looks like a straight line, and the slope of that line is v.

The **instantaneous speed** is simply the magnitude of the instantaneous velocity, $|v|$. So a particle with $v = -8\ \mathrm{m/s}$ has an instantaneous speed of $8\ \mathrm{m/s}$. Notice that the corresponding statement about *averages* is false: average speed is generally *not* the absolute value of average velocity, because the two are built from different quantities in the numerator (distance versus displacement). Our lap runner is the extreme case: average velocity zero, average speed nowhere near it.

Checkpoint 2.3

A car drives from one town to another. The trip takes $2\ \mathrm{h}$ and the towns are $120\ \mathrm{km}$ apart along the highway. The car's speedometer at one point during the trip reads $75\ \mathrm{km/h}$.

a. What is the car's average speed?
b. What does the speedometer reading represent?
c. Can the average speed and instantaneous speed ever be equal? Under what conditions?

Answer

Answers to Checkpoints

Checkpoint 2.3. (a) The average speed is $120\ \mathrm{km}/2\ \mathrm{h} = 60\ \mathrm{km/h}$: the distance divided by the time. (b) The speedometer shows the *instantaneous* speed: how fast the car is moving at that moment, not averaged over the trip. (c) Yes: they are equal throughout any interval in which the car moves at constant speed, and a car whose speed varies must still have instantaneous speed equal to its average speed at some moment (it cannot always be above the average or always below it).

This page titled 2.2: Speed and Velocity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.3: Acceleration

When a particle's velocity changes (whether in magnitude, direction, or both) the particle is said to be **accelerating**. The *average acceleration* over a finite time interval is the change in velocity divided by the time it took:

Key Equation: Average acceleration

$$a_{\text{avg}} = \frac{\Delta v}{\Delta t} = \frac{v_B - v_A}{t_B - t_A}, \tag{2.5}$$

and the *instantaneous acceleration* is again obtained by letting $\Delta t \to 0$:

Key Equation: Instantaneous acceleration

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}. \tag{2.6}$$

The second form says the same thing twice over: acceleration is the rate of change of velocity, and velocity is itself the rate of change of position, so acceleration is the rate of change of the rate of change of position. If that sounds like a mouthful, it is, which is exactly why we give it its own name and its own symbol.

What the sign of the acceleration does and does not tell you

Here is the single most common error in this chapter. It is tempting to read "negative acceleration" as "slowing down." It does not mean that. The sign of a tells you which way the *change* in velocity points, not whether the object is gaining or losing speed. To decide that, you must compare the sign of a with the sign of v:

Key Equation: Speeding up or slowing down

$$\textbf{Same signs } (v \text{ and } a \text{ both } + \text{ or both } -) \implies \textbf{speeding up.} \tag{2.3.1}$$

$$\textbf{Opposite signs} \implies \textbf{slowing down.} \tag{2.3.2}$$

All four combinations occur routinely:

Signs of velocity and acceleration: speeding up or slowing down

v	a	Example	Result
$+$	$+$	car moving right, pressing the accelerator	speeding up
$+$	$-$	car moving right, braking	slowing down
$-$	$-$	car moving left, pressing the accelerator	speeding up
$-$	$+$	car moving left, braking	slowing down

Note: On the word "deceleration"

Everyday speech uses *deceleration* to mean "slowing down," and you will see the word in this book and elsewhere. It is a statement about speed, not about sign: an object decelerating while moving in the $-x$ direction has a *positive* acceleration. When a problem gives you a deceleration as a bare positive number, you must supply the sign yourself by asking which way the object is moving. When in doubt, avoid the word and state the acceleration with its sign.

Example 2.1: Reading the signs

A train is moving in the $-x$ direction at $20\ \mathrm{m/s}$ when the engineer applies the brakes, producing a constant acceleration of $a = +2.5\ \mathrm{m/s^2}$ until the train stops.

a. Is the train speeding up or slowing down? How do you know?
b. How long does it take to stop?
c. After it stops, the same acceleration continues to act for another $2.0\ \mathrm{s}$. Describe the motion.

Solution

(a) The velocity is negative ($v_0 = -20\ \mathrm{m/s}$, since motion is in the $-x$ direction) while the acceleration is positive. Opposite signs, so the train is *slowing down*, even though the acceleration is positive.

(b) Because the acceleration is constant, the average acceleration of equation (2.5) equals a itself, so $a = (v - v_0)/\Delta t$. We want Δt, which sits in the denominator. Multiply both sides by Δt, then divide both sides by a:

$$a = \frac{v - v_0}{\Delta t} \quad \Longrightarrow \quad a\,\Delta t = v - v_0 \quad \Longrightarrow \quad \Delta t = \frac{v - v_0}{a}. \tag{2.3.3}$$

The train stops, so $v = 0$. Substitute, keeping the sign of v_0 inside parentheses:

$$\Delta t = \frac{0 - (-20\,\mathrm{m/s})}{2.5\,\mathrm{m/s^2}} = \frac{+20\,\mathrm{m/s}}{2.5\,\mathrm{m/s^2}} = 8.0\,\mathrm{s}. \tag{2.3.4}$$

Subtracting a negative number is the same as adding its magnitude, so the numerator is $+20$, not -20. Had we dropped the sign, the answer would have been $-8.0\ \mathrm{s}$: a time before the brakes were applied, which is a clear signal that a sign went astray. The units check too: $(\mathrm{m/s}) \div (\mathrm{m/s^2}) = (\mathrm{m/s}) \times (\mathrm{s^2/m}) = \mathrm{s}$.

(c) Nothing about the acceleration changes at the instant the train stops. Restart the clock there, so the new initial velocity is $v_0 = 0$. With $a > 0$ the velocity immediately becomes positive: the train begins moving in the $+x$ direction and speeds up. After two more seconds,

$$v = v_0 + a\,\Delta t = 0 + (2.5\,\mathrm{m/s^2})(2.0\,\mathrm{s}) = +5.0\,\mathrm{m/s}. \tag{2.3.5}$$

Velocity zero does not mean acceleration zero: a point we will meet again at the top of a thrown ball's flight.

Checkpoint 2.4

A ball rolls up a ramp, slows, stops, and rolls back down. Take "up the ramp" as the $+x$ direction. What is the sign of the ball's acceleration on the way up? At the instant it is momentarily at rest? On the way down? Does the sign ever change?

Answer

Answers to Checkpoints

Checkpoint 2.4. Negative in all three cases, so it never changes sign. On the way up the ball moves in $+x$ but is slowing, so its acceleration opposes its velocity and is negative. At the top the velocity is momentarily zero, but the velocity is still changing (from positive to negative), so the acceleration is still negative. On the way down the ball moves in $-x$ and speeds up, so acceleration and velocity share the same (negative) sign. The *velocity* changes sign at the top; the acceleration does not.

This page titled 2.3: Acceleration is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.4: Motion Graphs

Much of the physical content of this chapter can be read directly off a graph, without solving a single equation. Two ideas do all the work, and both follow from definitions you have already seen.

Slopes give you the next quantity down. Velocity is the derivative of position, and a derivative *is* a slope. So:

- The slope of a position–time graph at any instant is the **velocity** at that instant.
- The slope of a velocity–time graph at any instant is the **acceleration** at that instant.

Steep means fast; flat means stopped (on an x–t graph) or coasting at constant velocity (on a v–t graph); downward-sloping means negative.

Areas give you the previous quantity back. Running the derivative backwards, the area between a curve and the horizontal axis accumulates the quantity one level up:

- The area under a velocity–time graph over an interval is the **displacement** Δx during that interval.
- The area under an acceleration–time graph over an interval is the **change in velocity** Δv.

That the first of these is true is easiest to see for constant velocity: if v is a constant $5\ \mathrm{m/s}$ for $3\ \mathrm{s}$, the region under the graph is a rectangle of height 5 and width 3, and its "area" is $15\ \mathrm{m}$, exactly the displacement. Any varying velocity can be sliced into thin strips, each of which is nearly a rectangle of that kind, and the strips add up.

Key Equation: Areas under a velocity–time graph

$$\Delta x = \text{signed area under the } v\text{–}t \text{ curve}, \qquad \Delta v = \text{signed area under the } a\text{–}t \text{ curve}. \tag{2.7}$$

The word *signed* is essential. Area lying *above* the horizontal axis counts as positive; area lying *below* it counts as negative, and cancels against the positive part. This is exactly the distance-versus-displacement distinction wearing a different hat: add the signed areas and you get displacement; add their absolute values and you get distance traveled.

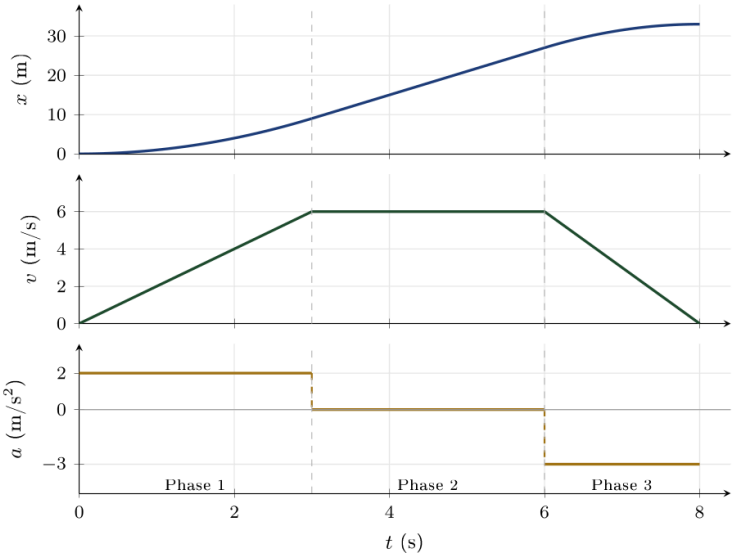


Figure 2.3. One motion, three graphs, one shared time axis. A car starts from rest and accelerates at $2\ \mathrm{m/s^2}$ for $3\ \mathrm{s}$ (Phase 1), cruises at $6\ \mathrm{m/s}$ for $3\ \mathrm{s}$ (Phase 2), then brakes at $3\ \mathrm{m/s^2}$ until it stops at $t = 8\ \mathrm{s}$ (Phase 3). The slope of the top graph is the middle graph; the slope of the middle graph is the bottom graph. Equivalently, the area under the bottom graph builds the middle one, and the area under the middle graph builds the top one.

•Figure description

Three stacked graphs sharing one time axis, describing a single motion of a car in three phases. The top graph shows position rising, first curving upward during phase one as the car accelerates from rest at two meters per second squared for three seconds, then rising as a straight line during phase two as it cruises at six meters per second for three seconds, then curving over as it brakes in phase three. The middle graph shows velocity: a straight rise to six meters per second, a horizontal segment, then a fall back to zero. The bottom graph shows acceleration as a step function: positive and constant, then zero, then negative and constant.

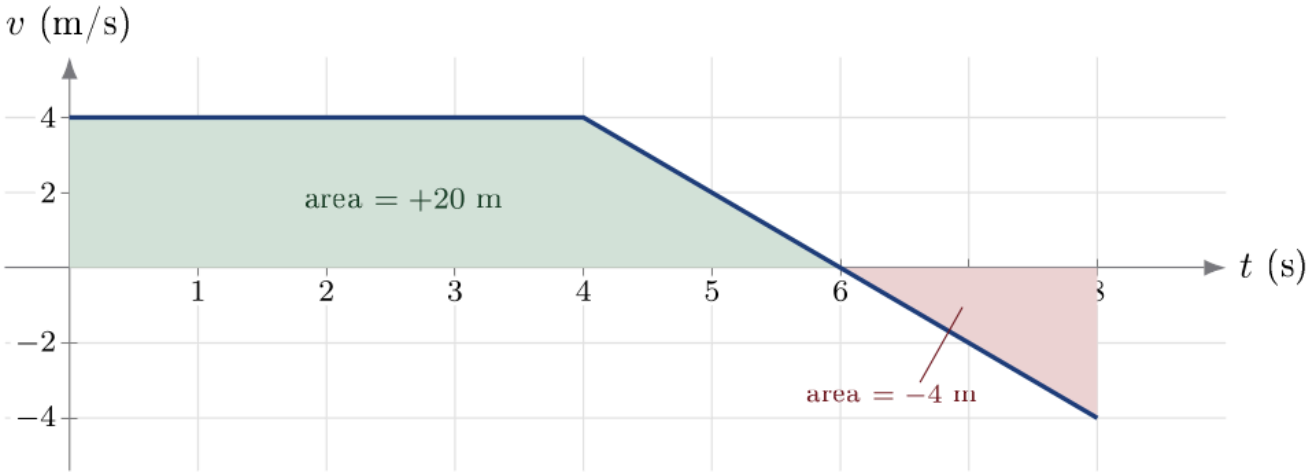


Figure 2.4. Displacement as signed area. Between $t = 0$ and $t = 6\ \mathrm{s}$ the velocity is positive and the shaded region above the axis contributes $+20\ \mathrm{m}$. Between $t = 6\ \mathrm{s}$ and $t = 8\ \mathrm{s}$ the velocity is negative and the region below the axis contributes $-4\ \mathrm{m}$. The net displacement over the full $8\ \mathrm{s}$ is $+16\ \mathrm{m}$; the total distance traveled is $24\ \mathrm{m}$.

•Figure description

A velocity-versus-time graph used to find displacement as signed area. From zero to six seconds the velocity is positive and the region between the curve and the time axis is shaded above the axis, contributing plus twenty meters. From six to eight seconds the velocity is negative and the shaded region lies below the axis, contributing a negative amount. The net displacement is the sum of the two signed areas.

Reading the three graphs together

The clearest way to see all of this at once is to stack the three graphs for a single motion with their time axes aligned, as in Figure 2.3. Trace a vertical line down through all three panels at any instant and read off position, velocity, and acceleration simultaneously. Notice in particular:

- Where a is constant and positive (Phase 1), v is a straight line with positive slope and x curves *upward*.
- Where $a = 0$ (Phase 2), v is flat and x is a straight line: constant velocity.
- Where a is constant and negative (Phase 3), v is a straight line with negative slope and x curves *downward*, flattening out as the car comes to rest.

The car is moving forward the entire time, so x never decreases, not even in Phase 3, where the acceleration is negative. Negative acceleration bends the position curve downward; it does not, by itself, send the car backwards.

Example 2.2: Getting everything from a velocity–time graph

Use the velocity–time graph in Figure 2.4 to find

a. the acceleration during each phase of the motion;
b. the time at which the particle reverses direction;
c. the displacement over the full $8\ \text{s}$;
d. the total distance traveled over the full $8\ \text{s}$;
e. the average velocity and average speed over the full $8\ \text{s}$.

Solution

(a) Acceleration is the slope of the v–t graph: rise over run, $\Delta v/\Delta t$. From $t = 0$ to $4\ \text{s}$ the graph is flat, so its rise is zero and $a = 0$. From $t = 4\ \text{s}$ to $8\ \text{s}$ the velocity falls from $+4$ to $-4\ \text{m/s}$. The change is always *final minus initial*:

$$a = \frac{\Delta v}{\Delta t} = \frac{(-4) - (+4)}{8 - 4}\ \frac{\text{m/s}}{\text{s}} = \frac{-8}{4}\ \text{m/s}^2 = -2.00\ \text{m/s}^2. \tag{2.4.1}$$

The velocity drops by $8\ \text{m/s}$, not $4\ \text{m/s}$, because it passes through zero on the way: 4 units down to the axis and 4 more below it.

(b) The particle reverses direction where the velocity changes sign, where the graph crosses the axis, at $t = 6\ \text{s}$. Note that this is *not* where the acceleration changes; the acceleration has been $-2\ \text{m/s}^2$ since $t = 4\ \text{s}$ and continues unchanged through the turnaround.

(c) Displacement is the signed area. On a v–t graph, a width is measured in seconds and a height in meters per second, so every area comes out in $\text{s} \times \text{m/s} = \text{m}$: a displacement, as it should be.

The region above the axis is a trapezoid (a four-sided figure with two parallel sides). Its parallel sides are horizontal: the top edge runs from $t = 0$ to $t = 4\ \text{s}$ (length $4\ \text{s}$), and the bottom edge along the axis runs from $t = 0$ to $t = 6\ \text{s}$ (length $6\ \text{s}$). The distance between them is the height, $4\ \text{m/s}$. The area of a trapezoid is the *average* of the parallel sides times the distance between them:

$$A_+ = \tfrac{1}{2}(4 + 6)(4)\ \text{m} = \tfrac{1}{2}(10)(4)\ \text{m} = +20.0\ \text{m}. \tag{2.4.2}$$

If you don't remember the trapezoid formula, split the region into a rectangle ($t = 0$ to $4\ \text{s}$) and a triangle ($t = 4\ \text{s}$ to $6\ \text{s}$): $(4)(4) + \tfrac{1}{2}(2)(4) = 16 + 4 = 20\ \text{m}$. Same answer.

The region below the axis is a triangle of base $2\ \text{s}$ and height $4\ \text{m/s}$. It lies below the axis, so its area counts as negative:

$$A_- = -\tfrac{1}{2}(2)(4)\ \text{m} = -4.00\ \text{m}. \tag{2.4.3}$$

The displacement is the sum of the signed areas, $\Delta x = A_+ + A_- = 20.0 + (-4.00) = +16.0\ \text{m}$.

(d) The distance ignores the signs and adds the magnitudes: $|A_+| + |A_-| = 20.0 + 4.00 = 24.0\ \text{m}$.

(e) Dividing each by the elapsed time,

$$v_{\text{avg}} = \frac{16.0\,\text{m}}{8.00\,\text{s}} = 2.00\,\text{m/s}, \qquad s_{\text{avg}} = \frac{24.0\,\text{m}}{8.00\,\text{s}} = 3.00\,\text{m/s}. \tag{2.4.4}$$

As promised, the average speed is not the absolute value of the average velocity.

Checkpoint 2.5

A position–time graph is a parabola opening *downward*, and the particle's position is increasing at $t = 0$. What are the signs of v and a at $t = 0$? Is the particle speeding up or slowing down? What eventually happens to it?

Answer

Answers to Checkpoints

Checkpoint 2.5. The position is increasing, so the slope of the x–t graph is positive: $v > 0$. A parabola opening downward curves downward, so $a < 0$. Opposite signs mean the particle is slowing down. It eventually comes momentarily to rest at the vertex of the parabola (its maximum x), then moves back in the $-x$ direction, speeding up as it goes, since then v and a are both negative.

This page titled 2.4: Motion Graphs is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.5: Kinematics in One Dimension

When a particle moves with *constant* acceleration a, the relationships among displacement Δx, initial velocity v_0, final velocity v, acceleration a, and elapsed time t simplify dramatically. Four equations result, each of which relates four of these five quantities.

Where the equations come from

Start from the definition of acceleration, $a = \frac{dv}{dt}$, and take a to be constant. Integrating both sides from time 0 to time t gives

$$\int_{v_0}^{v} dv' = \int_{0}^{t} a\, dt' \qquad \Longrightarrow \qquad v - v_0 = at, \tag{2.5.1}$$

which is the first equation. Now use the definition of velocity, $v = \frac{dx}{dt}$, substitute the result we just obtained, and integrate again:

$$\int_{x_0}^{x} dx' = \int_{0}^{t} (v_0 + at')\, dt' \qquad \Longrightarrow \qquad \Delta x = v_0 t + \tfrac{1}{2} a t^2, \tag{2.5.2}$$

the second equation. The remaining two are algebra: solve the first for t and substitute into the second to eliminate time, or average the first over the interval to eliminate the acceleration.

If integration is still new to you, Section 2.4 offers the same result without it. Under constant acceleration the v–t graph is a straight line from v_0 to v, and the region beneath it is a trapezoid of width t and parallel sides v_0 and v. Its area is $\frac{1}{2}(v_0 + v)t$, and by equation (2.7) that area *is* the displacement, which is the fourth equation, obtained by inspection.

Key Equation: The four kinematic equations

$$v = v_0 + a\,t \tag{2.8}$$
$$\Delta x = v_0\,t + \tfrac{1}{2} a\,t^2 \tag{2.9}$$
$$v^2 = v_0^2 + 2a\,\Delta x \tag{2.10}$$
$$\Delta x = \tfrac{1}{2}(v + v_0)\,t \tag{2.11}$$

Note: Two conditions on these four equations

First, **the acceleration must be constant** over the whole interval you apply them to. If the acceleration changes partway through, split the motion into segments on which it is constant and apply the equations to each segment in turn, carrying the final velocity of one segment in as the initial velocity of the next. If the acceleration varies *continuously*, these equations do not apply at all and you must integrate (see Example 2.5).

Second, t **here means the elapsed time** Δt, with the clock started at zero at the beginning of the interval. Likewise v_0 is the velocity at that starting instant and Δx is the displacement measured from the starting position, not the position itself.

Each equation is missing one of the five kinematic quantities, which makes it useful in different circumstances:

The constant-acceleration equations and when to use each

Equation	Missing variable	Useful when…
(2.8)	Δx	you want velocity at a given time.
(2.9)	v	you want position at a given time.
(2.10)	t	time is neither known nor wanted.
(2.11)	a	acceleration is neither known nor wanted.

Problem-solving strategy. For constant-acceleration problems:

1. Draw a diagram and choose a positive direction.
2. List the five quantities (Δx, v_0, v, a, t), marking which are known and which are unknown.
3. Identify the equation that includes the three knowns and the one unknown you want.
4. Solve algebraically before substituting numbers. Check units.

Example 2.3: Jet landing on an aircraft carrier

A jet lands on an aircraft carrier at a speed of $65\ \text{m/s}$.

a. What is the jet's acceleration (assumed constant) if it stops in $2.0\ \text{s}$ due to an arresting cable?
b. How many g's does the pilot experience, given that $1g = 9.8\ \text{m/s}^2$?
c. How far does the jet travel before coming to a stop?

Solution

Take the direction of the jet's motion as $+x$.

(a) Knowns: $v_0 = 65\ \text{m/s}$, $v = 0$, $t = 2.0\ \text{s}$. Unknown: a. Equation (2.8) contains exactly these quantities. Solve it for a before substituting: subtract v_0 from both sides, then divide both sides by t:

$$v = v_0 + at \quad \Longrightarrow \quad v - v_0 = at \quad \Longrightarrow \quad a = \frac{v - v_0}{t}. \tag{2.5.3}$$

Now substitute:

$$a = \frac{0 - 65\,\text{m/s}}{2.0\,\text{s}} = -32.5\,\text{m/s}^2. \tag{2.5.4}$$

The velocity is positive and the acceleration is negative (opposite signs), so the jet is slowing down, as it had better be.

(b) To express an acceleration "in g's," divide it by g. The units cancel, leaving a pure number that says how many times larger it is than free-fall acceleration:

$$\frac{|a|}{g} = \frac{32.5\ \cancel{\text{m/s}^2}}{9.8\ \cancel{\text{m/s}^2}} \approx 3.3. \tag{2.5.5}$$

The pilot experiences about $3.3\ g$: significant but survivable, and in the range that carrier arresting gear is designed to deliver.

(c) We could use equation (2.10), but notice that equation (2.11) needs only the two velocities and the time, all of them *given* data, none of them values we computed and rounded:

$$\Delta x = \tfrac{1}{2}(v + v_0)t = \tfrac{1}{2}(0 + 65\,\text{m/s})(2.0\,\text{s}) = 65\,\text{m}. \tag{2.5.6}$$

Here $\frac{1}{2}(v + v_0)$ is the average of the starting and ending velocities, $32.5\ \text{m/s}$, and average velocity times time is displacement. The jet travels $65\ \text{m}$ along the deck before stopping. Whenever you have a choice, prefer the equation built from given data over one built from your own intermediate results.

Example 2.4: Chasing down a fumble

A linebacker recovers a fumble and is running at $5.0\ \text{m/s}$ directly away from a wide receiver. If the receiver is initially $10\ \text{m}$ behind the linebacker, starts from rest, and accelerates uniformly at $2.5\ \text{m/s}^2$, how long does it take the receiver to catch the linebacker?

Solution

Step 1: Write a position equation for each player. Let $x = 0$ be the receiver's starting position, with positive x in the direction of motion, and start the clock when the chase begins. Both players obey $x = x_0 + v_0 t + \frac{1}{2}at^2$ (equation (2.9) with $\Delta x = x - x_0$); they differ only in their starting values.

The linebacker starts $10\ \text{m}$ ahead, so $x_0 = 10$, and runs at constant velocity, so $a = 0$ and the t^2 term vanishes:

$$x_L(t) = 10 + 5.0\,t. \tag{2.5.7}$$

The receiver starts at the origin ($x_0 = 0$) from rest ($v_0 = 0$), so only the acceleration term survives:

$$x_R(t) = \tfrac{1}{2}(2.5)t^2 = 1.25\,t^2. \tag{2.5.8}$$

(Here x is in meters and t in seconds.)

Step 2: Set the positions equal. "Catches" means both players are at the same place at the same time, $x_R(t) = x_L(t)$:

$$1.25\,t^2 = 10 + 5.0\,t. \tag{2.5.9}$$

Step 3: Put the equation in standard quadratic form. The equation contains t^2, so it is quadratic. To use the quadratic formula it must read $at^2 + bt + c = 0$, with everything on one side. Subtract 10 and subtract $5.0\ t$ from both sides:

$$1.25\,t^2 - 5.0\,t - 10 = 0. \tag{2.5.10}$$

Reading off the coefficients, *signs included*: $a = 1.25$, $b = -5.0$, $c = -10$. (This a is just the coefficient in the quadratic, not the acceleration.)

Step 4: Apply the quadratic formula. The formula is

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{2.5.11}$$

Work out each piece separately, since this is where sign errors happen:

$$\begin{aligned} -b &= -(-5.0) = +5.0, & b^2 &= (-5.0)^2 = 25, \\ -4ac &= -4(1.25)(-10) = +50, & 2a &= 2(1.25) = 2.5. \end{aligned} \tag{2.5.12}$$

The two negatives in $-4ac$ multiply to a positive, so the quantity under the square root is $25 + 50 = 75$:

$$t = \frac{5.0 \pm \sqrt{75}}{2.5} = \frac{5.0 \pm 8.66}{2.5}. \tag{2.5.13}$$

Step 5: Choose the physical root. A quadratic has two roots, and physics has to pick. The "minus" root gives $t = (5.0 - 8.66)/2.5 = -1.46\ \text{s}$, a time before the chase began, so it is discarded. The "plus" root gives

$$t = \frac{5.0 + 8.66}{2.5} = \frac{13.66}{2.5} \approx 5.46\,\text{s}, \tag{2.5.14}$$

so the receiver catches the linebacker in about $\boxed{5.5\,\text{s}}$.

Check. Put $t = 5.46\ \text{s}$ back into both position equations. Receiver: $1.25(5.46)^2 = 37.3\ \text{m}$. Linebacker: $10 + 5.0(5.46) = 37.3\ \text{m}$. They agree, so the players are indeed side by side.

Checkpoint 2.6

In Example 2.4, how fast is the receiver moving at the instant of catching the linebacker? Is the receiver faster or slower than the linebacker at that moment? Does your answer seem physically reasonable for a human sprinter? (For reference, the fastest speed ever recorded for a human runner is about $12.4\ \text{m/s}$.)

Answer

Example 2.5: When the four equations do not apply

A cart is pushed along the x-axis so that its acceleration increases steadily with time: $a(t) = 6t$, with a in m/s^2 and t in seconds. It starts from rest at the origin. Find its velocity and position at $t = 2.0\ \text{s}$.

Solution

The acceleration is not constant, so equations (2.8)–(2.11) are simply unavailable. We must go back to the definitions and integrate, as in the derivation of Section 2.5, but this time with the actual $a(t)$ in place of a constant.

Step 1: Velocity. Since $a = \frac{dv}{dt}$, the change in velocity from time 0 to time t is the integral of a over that interval:

$$v(t) - v(0) = \int_0^t a(t')\,dt' = \int_0^t 6t'\,dt'. \tag{2.5.15}$$

(The prime on t' just distinguishes the variable being integrated over from the upper limit t.) The cart starts from rest, so $v(0) = 0$. To integrate, use the power rule: raise the exponent by one and divide by the new exponent, $\int t^n\ dt = t^{n+1}/(n+1)$. Here $n = 1$:

$$\int_0^t 6t'\,dt' = 6\left[\frac{t'^2}{2}\right]_0^t = 6\left(\frac{t^2}{2} - \frac{0^2}{2}\right) = 3t^2. \tag{2.5.16}$$

So $v(t) = 3t^2$, and at $t = 2.0\ \text{s}$, $v = 3(2.0)^2 = 3(4.0) = 12.0\ \text{m/s}$.

Step 2: Position. Since $v = \frac{dx}{dt}$, integrate once more. The cart starts at the origin, so $x(0) = 0$. Now $n = 2$:

$$x(t) = \int_0^t 3t'^2\,dt' = 3\left[\frac{t'^3}{3}\right]_0^t = 3\left(\frac{t^3}{3} - 0\right) = t^3, \tag{2.5.17}$$

so $x = (2.0)^3 = 8.00\ \text{m}$.

Check. Differentiate back down: $\frac{d}{dt}(t^3) = 3t^2$, which is v, and $\frac{d}{dt}(3t^2) = 6t$, which is a. Each integration undid one differentiation, as it should.

It is worth seeing how badly the constant-acceleration equations would have failed here. At $t = 2.0\ \text{s}$ the acceleration happens to be $a = 6(2.0) = 12\ \text{m/s}^2$; feeding that into equation (2.9) as though it had held all along gives $\Delta x = \frac{1}{2}(12)(2.0)^2 = 24\ \text{m}$, three times the correct answer. The equations are not approximations that degrade gracefully when the acceleration varies: they are derived from an assumption, and when the assumption fails, so do they.

Answers to Checkpoints

Checkpoint 2.6. Starting from rest with $a = 2.5\ \text{m/s}^2$, after $t = 5.46\ \text{s}$ the receiver moves at $v = at = (2.5)(5.46) \approx 13.7\ \text{m/s}$, well over twice the linebacker's $5.0\ \text{m/s}$, as it must be for the receiver to have closed a $10\ \text{m}$ gap. It is *not* physically reasonable: $13.7\ \text{m/s}$ exceeds the fastest human speed ever recorded, and no one can keep accelerating uniformly for five and a half seconds. The constant-acceleration model is an idealization that breaks down here; a real receiver would reach top speed sooner and take longer to catch up.

This page titled 2.5: Kinematics in One Dimension is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.6: Free Fall

A particularly important case of constant acceleration is **free fall**: motion under the influence of gravity alone, with air resistance neglected. Near Earth's surface, all objects in free fall accelerate downward at the same rate, regardless of mass:

Key Equation: Free-fall acceleration near Earth's surface

$$g = 9.8\,\mathrm{m/s^2} \quad \text{(downward)}. \tag{2.12}$$

Note: The sign of g — read this twice

The symbol g denotes a *positive* number, $9.8\ \mathrm{m/s^2}$. It is the *magnitude* of the free-fall acceleration, and it never carries a sign of its own. The direction is supplied separately, by you, when you choose an axis:

- If you take **up** as positive (the usual choice), then $a = -g = -9.8\ \mathrm{m/s^2}$.
- If you take **down** as positive, then $a = +g = +9.8\ \mathrm{m/s^2}$.

Writing "$g = -9.8$" and then also writing "$a = -g$" is a double negative that will silently flip your answer. Decide on the axis first, write a with the correct sign once, and never touch the sign again.

The name "free fall" is misleading in one important respect: an object does not have to be falling to be in free fall. The term describes any motion in which gravity is the only force acting, so a ball thrown straight up is in free fall from the instant it leaves your hand, on the way up, at the peak, and on the way down alike. Its acceleration is g in the downward direction throughout, without interruption. In particular, at the highest point of its flight the ball's *velocity* is momentarily zero, but its *acceleration* is still $9.8\ \mathrm{m/s^2}$ downward. If the acceleration really did vanish there, the ball would simply hang in the air.

This universality of free fall was argued for by Galileo in the early seventeenth century, on the basis of careful experiments with balls rolling down inclined planes (which slow the motion enough to time it) rather than the often-repeated but almost certainly apocryphal drop from the Leaning Tower of Pisa. It was demonstrated far more dramatically in 1971, when Apollo 15 commander David Scott stood on the Moon and released a geological hammer and a falcon feather at the same instant. They struck the lunar surface together, because the Moon has no atmosphere and thus no air resistance. Free-fall acceleration at the Moon's surface is only about one-sixth of its value at Earth's, a difference explored in Chapter 13.

Two caveats are worth keeping in view. First, g is not truly a universal constant even on Earth: it ranges from roughly $9.78\ \mathrm{m/s^2}$ at the equator to $9.83\ \mathrm{m/s^2}$ at the poles, and falls off slowly with altitude. The standard value is $9.80665\ \mathrm{m/s^2}$; we will use $9.8\ \mathrm{m/s^2}$ throughout. Second, neglecting air resistance is an approximation, and how good it is depends on the object. For a dense, compact object falling a short distance (a dropped wrench, a bumped volleyball) it is excellent. For a feather, a sheet of paper, a parachutist, or anything falling long enough to approach terminal speed, it is not, and the results of this section will not describe the motion. We will learn how to handle air resistance properly in Chapter 6.

Example 2.6: Pinecone from a tall tree

A pinecone falls from rest from the top of a $25\ \mathrm{m}$ tall tree.

a. How long does it take to reach the ground?
b. How fast is it moving when it hits?

Solution

Choose upward as positive, with the origin at the ground. Then $y_0 = 25\ \mathrm{m}$, $v_0 = 0$, and $a = -g = -9.8\ \mathrm{m/s^2}$. The pinecone reaches the ground when $y = 0$, so its displacement is final minus initial, $\Delta y = 0 - 25\ \mathrm{m} = -25\ \mathrm{m}$.

(a) Knowns: Δy, v_0, a. Unknown: t. Equation (2.9) contains exactly these:

$$\Delta y = v_0 t + \tfrac{1}{2} a t^2 \quad\Longrightarrow\quad -25 = 0 + \tfrac{1}{2}(-9.8)t^2 = -4.9\,t^2. \tag{2.6.1}$$

The $v_0 t$ term is zero because the pinecone starts from rest. To isolate t^2, divide both sides by -4.9:

$$t^2 = \frac{-25}{-4.9} = 5.10\,\text{s}^2. \tag{2.6.2}$$

The two minus signs cancel, as they must: a negative displacement and a negative acceleration are consistent with a real, positive elapsed time. (Had the signs disagreed, we would be asked for the square root of a negative number, a sure sign of a sign error.) Take the square root of both sides, keeping the positive root because elapsed time is positive:

$$t = \sqrt{5.10\,\text{s}^2} \approx 2.26\,\text{s}. \tag{2.6.3}$$

(b) Use equation (2.8):

$$v = v_0 + at = 0 + (-9.8\,\text{m/s}^2)(2.26\,\text{s}) \approx -22.1\,\text{m/s}. \tag{2.6.4}$$

The negative sign indicates downward motion; the pinecone strikes the ground at a speed of about $22\ \text{m/s}$.

Check. Equation (2.10) gives the final speed from the given data alone, without the rounded time: $v^2 = v_0^2 + 2a\ \Delta y = 0 + 2(-9.8)(-25) = 490\ \text{m}^2/\text{s}^2$, so $|v| = \sqrt{490} = 22.1\ \text{m/s}$. The square root returns only the size; we supply the minus sign ourselves because the pinecone is moving down.

Example 2.7: Volleyball served straight up

A volleyball is served straight up from a height of $1.4\ \text{m}$ above the ground with an initial vertical speed of $11\ \text{m/s}$. What is the maximum height the ball reaches above the ground?

Solution

At the peak of its trajectory the ball's instantaneous velocity is zero: this is the condition that locates the highest point, and it is what makes the problem solvable. Take upward as positive, with $v_0 = 11\ \text{m/s}$, $v = 0$, and $a = -g = -9.8\ \text{m/s}^2$. The time is neither known nor wanted, so use equation (2.10). Solve it for Δy first: subtract v_0^2 from both sides, then divide both sides by $2a$:

$$v^2 = v_0^2 + 2a\Delta y \quad\Longrightarrow\quad v^2 - v_0^2 = 2a\,\Delta y \quad\Longrightarrow\quad \Delta y = \frac{v^2 - v_0^2}{2a}. \tag{2.6.5}$$

Substitute:

$$\Delta y = \frac{0 - (11\,\text{m/s})^2}{2(-9.8\,\text{m/s}^2)} = \frac{-121\,\text{m}^2/\text{s}^2}{-19.6\,\text{m/s}^2} \approx 6.17\,\text{m}. \tag{2.6.6}$$

Numerator and denominator are both negative, so their ratio is positive: the ball rises, as it should. The units reduce to meters because $(\text{m}^2/\text{s}^2) \div (\text{m/s}^2) = \text{m}$.

This is the height *above the point of release*, not above the ground. The maximum height above the ground is

$$y_{\max} = 1.4\,\text{m} + 6.17\,\text{m} = 7.57\,\text{m} \approx 7.6\,\text{m}, \tag{2.6.7}$$

rounded to one decimal place because $1.4\ \text{m}$ is known only to the nearest tenth (the addition rule of Chapter 1).

Note that the acceleration is $-9.8\ \text{m/s}^2$ at every instant of this calculation, including the peak itself.

Checkpoint 2.7

In Example 2.7, how long after being served does the volleyball return to its starting height of $1.4\ \text{m}$? *(Hint: by symmetry, the time to rise equals the time to fall back.)* What is its velocity at that moment, and how does it compare to the serve velocity?

Answer

Answers to Checkpoints

Checkpoint 2.7. At the peak $v = 0$, so the rise takes $t = v_0/g = (11)/(9.8) \approx 1.12$ s. By symmetry the fall back to the release height takes the same time, so the ball returns to 1.4 m about $2(1.12) \approx 2.2$ s after the serve. Its velocity then is $v = v_0 - gt = 11 - (9.8)(2.24) = -11$ m/s: the same speed as the serve, but directed downward. (Equation (2.10) gives this directly: returning to $\Delta y = 0$ means $v^2 = v_0^2$.)

This page titled 2.6: Free Fall is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.S: Motion in One Dimension (Summary)

Summary

- In one dimension, x, v, and a are signed scalars: the **sign is the direction**. Choose an origin and a positive direction at the start of a problem and keep them fixed.
- **Position** x is a coordinate; **displacement** $\Delta x = x_B - x_A$ is a change in position whose sign indicates direction; **distance** is the total path length and is never negative. The two differ exactly when the object reverses direction.
- **Average velocity**: $v_{\text{avg}} = \Delta x/\Delta t$. **Instantaneous velocity**: $v = \mathrm{d}x/\mathrm{d}t$. **Speed** is the magnitude of velocity; **average speed** uses distance, not displacement, and is therefore not generally $|v_{\text{avg}}|$.
- **Average acceleration**: $a_{\text{avg}} = \Delta v/\Delta t$. **Instantaneous acceleration**: $a = \mathrm{d}v/\mathrm{d}t = \mathrm{d}^2x/\mathrm{d}t^2$.
- The sign of a does not tell you whether an object is speeding up. Compare it to the sign of v: **same signs** $\Rightarrow$ **speeding up, opposite signs** $\Rightarrow$ **slowing down.**
- On graphs: the **slope** of x–t is v and the slope of v–t is a; the **signed area** under v–t is Δx and the signed area under a–t is Δv. Sum the signed areas for displacement, their absolute values for distance.
- For **constant acceleration**, the four kinematic equations (2.8)–(2.11) relate Δx, v_0, v, a, and t; each equation omits exactly one of the five quantities. They fail outright if the acceleration is not constant: integrate instead.
- Near Earth's surface, all objects in **free fall** accelerate downward at $g = 9.8\ \mathrm{m/s^2}$, independent of mass. Here g is a positive number: with up chosen as positive, $a = -g$. An object thrown upward is in free fall the whole time, including at the peak, where $v = 0$ but the acceleration is still g downward.

Key Equations

Displacement

$$\Delta x = x_B - x_A \tag{2.S.1}$$

Average velocity

$$v_{\text{avg}} = \frac{\Delta x}{\Delta t} = \frac{x_B - x_A}{t_B - t_A} \tag{2.S.2}$$

Instantaneous velocity

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{\mathrm{d}x}{\mathrm{d}t} \tag{2.S.3}$$

Average acceleration

$$a_{\text{avg}} = \frac{\Delta v}{\Delta t} = \frac{v_B - v_A}{t_B - t_A} \tag{2.S.4}$$

Instantaneous acceleration

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2x}{\mathrm{d}t^2} \tag{2.S.5}$$

Speeding up or slowing down

- Same signs of v and a $\Longrightarrow$ speeding up.
- Opposite signs $\Longrightarrow$ slowing down.

Areas under a velocity–time graph

$$\Delta x = \text{signed area under the } v\text{--}t \text{ curve}, \qquad \Delta v = \text{signed area under the } a\text{--}t \text{ curve} \tag{2.S.6}$$

The four kinematic equations (constant a)

$$\begin{aligned} v &= v_0 + a\,t \\ \Delta x &= v_0\,t + \tfrac{1}{2} a\,t^2 \\ v^2 &= v_0^2 + 2a\,\Delta x \\ \Delta x &= \tfrac{1}{2}(v + v_0)\,t \end{aligned} \tag{2.S.7}$$

Free-fall acceleration near Earth's surface

$$g = 9.8\,\mathrm{m/s^2} \quad \text{(downward)} \tag{2.S.8}$$

This page titled 2.S: Motion in One Dimension (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

2.E: Motion in One Dimension (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Problems tagged *[Graph-reading]* draw on Section 2.4 and are distributed among the groups below rather than collected into one.

Sections 2.1–2.2 — Position, Displacement, Velocity, and Speed

2.1 ★ A car travels east at $25\ \text{m/s}$ for $10\ \text{s}$, then west at $15\ \text{m/s}$ for $20\ \text{s}$.

a. What is the car's displacement?
b. What is the total distance traveled?
c. What is the average velocity?
d. What is the average speed?

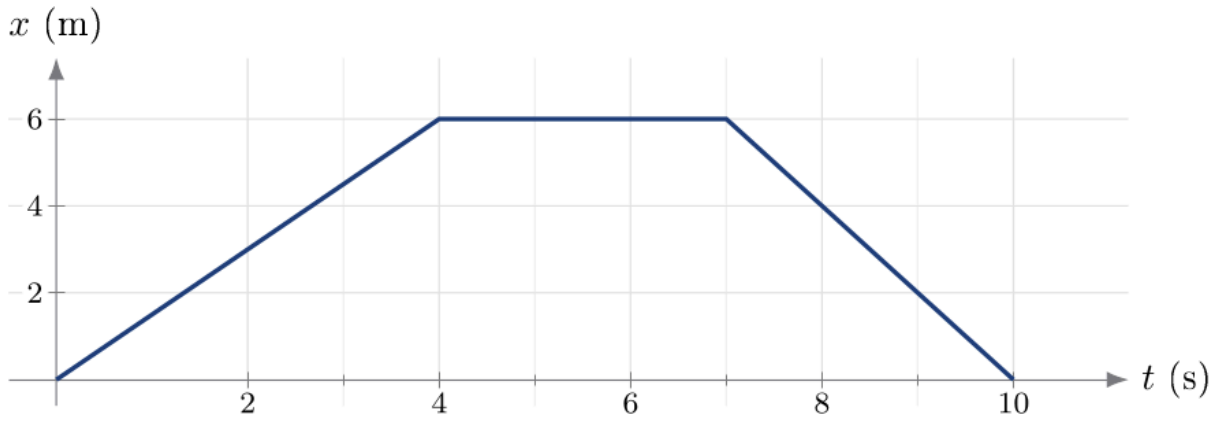


Figure 2.5. Position vs. time for Problem 2.2.

•Figure description

A position-versus-time graph made of three straight segments. Position rises steadily from zero at time zero to six meters at four seconds, stays constant at six meters from four to seven seconds, then falls steadily back to zero at ten seconds.

2.2 ★ *[Graph-reading.]* A particle's position along the x-axis as a function of time is shown in Figure 2.5.

a. Find the displacement over $t \in [0, 10]\ \text{s}$.
b. Find the total distance traveled over $t \in [0, 10]\ \text{s}$.
c. Find the average velocity and average speed over $t \in [0, 10]\ \text{s}$.
d. During which interval(s) is the particle at rest?

2.3 ★★ *[Real-world: sports.]* A swimmer completes one length of a $50\ \text{m}$ pool in $24\ \text{s}$, rests for $6.0\ \text{s}$, then swims back to the starting end in $28\ \text{s}$.

a. What is the total distance traveled?
b. What is the total displacement?
c. What is the average speed for the entire workout (including the rest)?
d. What is the average velocity for the entire workout?
e. Compute the average speed for each lap separately. Which lap was faster?

2.4 ★ *[Conceptual.]* Is it possible for an object to have a nonzero average speed but zero average velocity over the same time interval? If so, describe a specific situation in which this happens; if not, explain why not.

2.5 ★★ A particle moves along the x-axis according to $x(t) = 3t^2 - 2t + 5$, with x in meters and t in seconds.

a. Find expressions for $v(t)$ and $a(t)$.
b. What is the particle's velocity at $t = 2$ s?
c. Is the particle ever momentarily at rest? If so, when?
d. Find the average velocity over the interval $t \in [0, 2]$ s.
e. Compare your answer in (d) to the instantaneous velocity at $t = 1$ s. Why are they equal?

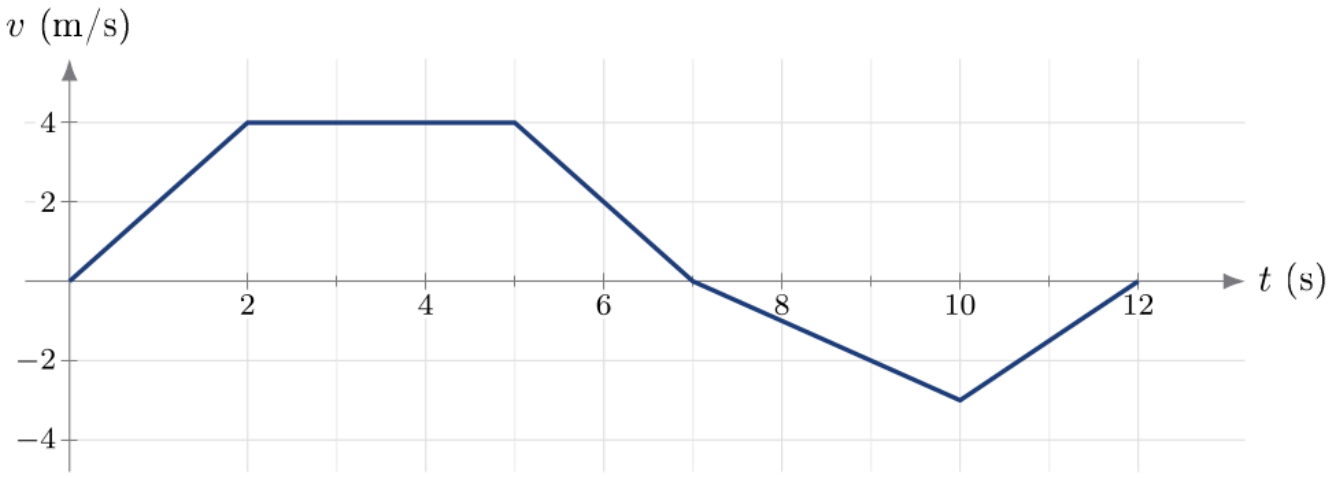


Figure 2.6. Velocity vs. time for Problem 2.6.

•Figure description

A velocity-versus-time graph made of straight segments. Velocity rises from zero to four meters per second between zero and two seconds, holds at four meters per second until five seconds, falls through zero at seven seconds, continues down to negative three meters per second at ten seconds, then returns to zero at twelve seconds.

2.6 ★★ *[Graph-reading.]* Figure 2.6 shows the velocity of a particle moving along the x-axis.

a. During which time interval(s) is the particle speeding up?
b. During which time interval(s) is the particle slowing down?
c. At what time(s) is the particle momentarily at rest?
d. Estimate the displacement of the particle over $t \in [0, 7]$ s from the area under the curve.
e. Sketch the corresponding acceleration–time graph.

2.7 ★ *[Conceptual.]* A car's speedometer reads 60 mi/hr. Explain why this reading alone does not fully specify the car's velocity. Give an example of two situations with the same speedometer reading but different velocities.

Section 2.3 — Acceleration

2.8 ★ A car accelerates uniformly from rest to 27 m/s in 9.0 s.

a. What is its acceleration?
b. How far does it travel during this time?
c. What is its average velocity over this interval?

2.9 ★★ *[Real-world: engineering.]* A passenger elevator in a tall building accelerates from rest at 1.2 m/s^2 for 2.0 s, cruises at constant velocity for 8.0 s, then slows uniformly to a stop over 3.0 s.

a. What is the cruise (top) speed?
b. What is the acceleration during the final segment? Give its sign, taking upward as positive and assuming the elevator is ascending.
c. How far does the elevator travel during each of the three segments?
d. What is the total distance traveled?
e. Sketch the velocity–time graph for the complete trip, and verify that the total area under it matches your answer to (d).

2.10 ★ *[Conceptual.]* "An object that is slowing down has a negative acceleration." Is this statement always true? Explain, and give an example of an object that is slowing down while having a *positive* acceleration. *(Hint: consider the direction of motion.)*

Section 2.5 — Kinematics in One Dimension

2.11 ★★ A speeding motorist passes a parked police car at a constant $30\ \text{m/s}$. At the instant the speeder passes, the police car begins accelerating from rest at $3.0\ \text{m/s}^2$ in pursuit.

a. How long does it take the police car to catch up?
b. How far has each car traveled at that moment?
c. How fast is the police car traveling when it catches up?
d. By how much does the police car's speed exceed the speeder's at that instant?
e. Convert your answer to (c) into mi/hr and comment on whether the pursuit, as modeled, is realistic.

2.12 ★★ *[Real-world: sports.]* A drag racer's car can accelerate uniformly from rest at $12\ \text{m/s}^2$. A standard quarter-mile drag strip is $402\ \text{m}$ long.

a. How long does it take to cover the quarter-mile, assuming the acceleration is maintained?
b. What is the speed at the end of the quarter-mile? Express in both m/s and mi/hr.
c. Real top-fuel dragsters, back when they raced the full quarter-mile, covered it in about $4.4\ \text{s}$ at exit speeds near $150\ \text{m/s}$. Compare your answers in (a) and (b) to these values; what does the discrepancy tell you about the assumption of constant acceleration?

2.13 ★★ *[Conceptual.]* Two cars start side by side. Car A accelerates uniformly from rest at $2.0\ \text{m/s}^2$. Car B starts $4.0\ \text{s}$ later, also from rest, accelerating at $3.0\ \text{m/s}^2$.

a. Will Car B catch up to Car A? If so, when (measured from Car A's start)?
b. How far have they each traveled at that point?
c. Sketch the position–time graphs of both cars on the same axes.

Section 2.6 — Free Fall

2.14 ★ A ball is thrown straight upward from ground level with an initial speed of $15\ \text{m/s}$.

a. How high does it rise?
b. How long is it in the air before returning to ground level?
c. With what velocity does it strike the ground?

2.15 ★★ A juggler tosses a ball straight up with an initial speed of $6.0\ \text{m/s}$ from a height of $1.2\ \text{m}$ above the ground.

a. How long is the ball in the air?
b. What is its velocity just before it hits the ground?
c. What is the maximum height above the ground?
d. At what time(s) is the ball at a height of $2.5\ \text{m}$ above the ground?

2.16 ★★ *[Real-world: biology.]* A flea can launch itself nearly straight up with a takeoff speed of about $1.3\ \text{m/s}$, lasting for a takeoff time of roughly $1.0\ \text{ms}$.

a. What is the flea's acceleration during the takeoff phase? Express in units of g.
b. Neglecting air resistance, how high above its starting height does the flea rise after lift-off?
c. How long is the flea in the air after lift-off (rise + fall back to starting height)?
d. Comment on whether neglecting air resistance is a good approximation for a flea (no calculation required).

2.17 ★★★ A stone is dropped from rest from the top of a cliff. A second stone is thrown straight down $1.0\ \text{s}$ later with an initial speed of $20\ \text{m/s}$. Both stones hit the ground simultaneously.

a. How long is the first stone in the air?
b. How tall is the cliff?

c. What is the speed of each stone just before it hits the ground?

2.18 ★★★ *[Real-world: engineering.]* A construction worker drops a wrench from the top of a building. An observer at a window $1.5\ \text{m}$ below the roof line hears the impact $4.0\ \text{s}$ after the wrench passes the window. For parts (a)–(c), assume free fall and ignore the time for sound to travel from the ground back up to the observer.

a. What is the speed of the wrench when it passes the observer's window?
b. How tall is the building (i.e., how far is the ground below the roof)?
c. What is the speed of the wrench at impact?
d. Repeat (b), but now account for the time it takes the sound of impact to travel back up to the observer at $v_{\text{sound}} = 343\ \text{m/s}$. Is the correction significant?

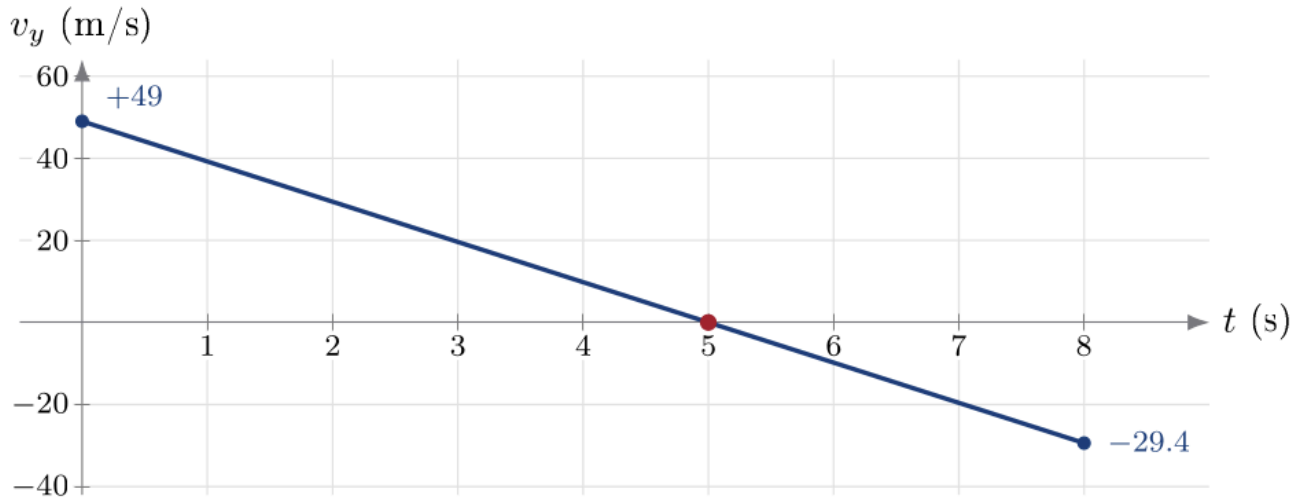


Figure 2.7. Velocity vs. time for Problem 2.19. The red dot marks the instant at which $v_y = 0$.

•Figure description

A velocity-versus-time graph for free fall: a single straight line of constant negative slope, starting at positive forty-nine meters per second at time zero and falling to negative twenty-nine point four meters per second at eight seconds; both endpoint values are labeled. A red dot marks the instant at five seconds where the line crosses zero and the vertical velocity is momentarily zero.

2.19 ★★★ *[Graph-reading.]* A particle moves along the vertical y-axis under the influence of gravity alone. The graph in Figure 2.7 shows its velocity as a function of time after it is launched from $y = 0$ at $t = 0$. Take upward as positive.

a. What is the initial velocity of the particle?
b. Read off the slope of the line. Is this consistent with motion under gravity alone near Earth's surface? Explain.
c. At what time does the particle reach its highest point, and how do you know from the graph?
d. Using the area under the graph, find the maximum height reached above the launch point.
e. At $t = 8\ \text{s}$, is the particle above, at, or below its starting height? Justify your answer using areas.
f. Sketch the corresponding position–time graph for $0 \le t \le 8\ \text{s}$.

2.20 ★★ *[Conceptual.]* A ball is thrown straight up into the air. At the very top of its flight, it is momentarily at rest.

a. What is the ball's velocity at the highest point?
b. What is the ball's acceleration at the highest point? *(Be careful: this is the part students most often get wrong.)*
c. A classmate argues that "if the velocity is zero, the acceleration must be zero too." Explain why this reasoning is incorrect.
d. If the acceleration really were zero at the top, describe what the ball would do next.

Challenge Problems

2.21 ★★★ *[Real-world: engineering.]* A model rocket lifts off from rest and burns its motor for $4.0\ \text{s}$, during which it has a constant upward acceleration of $30\ \text{m/s}^2$. At burnout the motor cuts off and the rocket coasts upward, then falls back; treat the post-burnout motion as free fall and neglect air resistance.

a. Find the rocket's velocity and altitude at the instant of burnout.
b. How much additional height does it gain while coasting to its peak?
c. What is its maximum altitude above the launch pad?
d. What is the total time from liftoff until it returns to the ground?
e. How fast is it moving just before it strikes the ground?

2.22 ★★★ A particle moves along the x-axis with velocity $v(t) = 3t^2 - 12t + 9$, where v is in $\mathrm{m/s}$ and t in seconds, and it starts at the origin, $x(0) = 0$.

a. Find the acceleration $a(t)$.
b. At what time(s) in $[0, 4]\ \mathrm{s}$ is the particle momentarily at rest?
c. During which subintervals is it moving in the $+x$ direction, and during which in the $-x$ direction?
d. Find the particle's displacement over $[0, 4]\ \mathrm{s}$.
e. Find the total distance it travels over $[0, 4]\ \mathrm{s}$, and explain why this differs from (d).
f. Explain why the four kinematic equations of Section 2.5 cannot be used for any part of this problem.

2.23 ★★★ *[Real-world: forensics.]* A flowerpot falls from rest off a high balcony. A resident on a lower floor sees it pass their window, which is $1.5\ \mathrm{m}$ tall, and times the pass at $0.20\ \mathrm{s}$.

a. How fast is the pot moving as it reaches the top of the window?
b. How fast is it moving at the bottom of the window?
c. From what height above the top of the window was the pot dropped?
d. How long has the pot been falling by the time it reaches the bottom of the window?

This page titled 2.E: Motion in One Dimension (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3: VECTORS

In Chapter 2 we restricted ourselves to motion along a single straight line. That simplification let us describe position, velocity, and acceleration using magnitudes and a sign. Most of the world is not so kind: a thrown football, a wind blowing across a runway, the pull of a rope on a sled. All of these have not just a magnitude but a *direction*, and any description that ignores direction is incomplete. The mathematical tool that captures both magnitude and direction is the **vector**.

This chapter develops the vector techniques we will use throughout the rest of the book. By the end of it, the next chapter's analysis of two-dimensional motion will look almost identical to the one-dimensional kinematics of Chapter 2; the only change is that scalars become vectors, and ordinary algebra becomes vector algebra.

Learning Objectives

- Distinguish scalar quantities from vector quantities and recognize examples of each.
- Resolve a vector into its x, y, and z components, and assign each component the correct sign.
- Write a vector in component, angle–magnitude, and unit-vector forms.
- Recover a vector's magnitude and direction from its components, placing the angle in the correct quadrant.
- Find the unit vector in the direction of any nonzero vector.
- Add, subtract, and scale vectors using their components.
- Compute the dot product and the cross product of two vectors, and interpret each geometrically.
- Use the dot product to find the angle between two vectors and to project one vector onto another.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

3.1: Scalars and Vectors
3.2: Components of a Vector
3.3: Unit Vectors
3.4: Vector Addition
3.5: Vector Multiplication
3.S: Vectors (Summary)
3.E: Vectors (Problems)

This page titled 3: Vectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.1: Scalars and Vectors

A **scalar** is a quantity that has only magnitude: a single number, which may be positive, negative, or zero, and which usually carries units. Mass, time, temperature, energy, and electric charge are all scalars; their value is fully specified by giving one number. A few scalars, such as a count of objects or a refractive index, are dimensionless and carry no units at all. Scalars live on a number line.

A **vector** is a quantity that has both magnitude and direction. Displacement, velocity, acceleration, force, and momentum are all vectors. The shortest description of a vector is "a scalar with a direction attached to it," and that intuition is correct: the magnitude of a vector is itself a scalar quantity (always nonnegative) representing the vector's length, while the direction tells you which way the vector points.

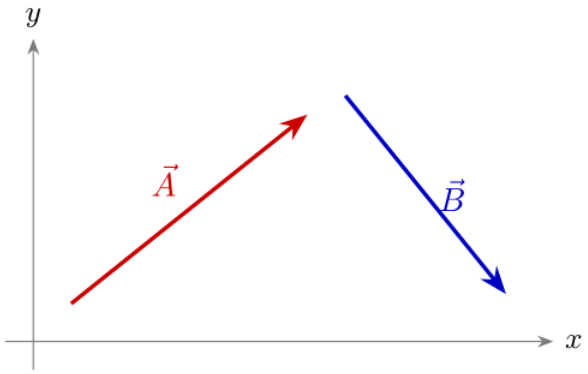


Figure 3.1. Two vectors of equal magnitude but different directions are *not* equal. Here $|\mathbf{A}| = |\mathbf{B}|$, yet $\mathbf{A} \neq \mathbf{B}$, because a vector is specified by magnitude *and* direction together.

•Figure description

Two arrows drawn from the origin on x and y axes. They have the same length but point in different directions, illustrating that equal magnitude alone does not make two vectors equal.

We denote a vector by a boldface symbol or with an arrow above it. In handwriting the arrow is universal; in print, both styles are common. Throughout this book we use the arrow notation: $\mathbf{A}$, $\mathbf{v}$, $\mathbf{F}$. The **magnitude** of a vector is written either as $|\mathbf{A}|$ or simply as the same symbol without the arrow: A. Two vectors are equal only if they agree in *both* respects: same magnitude *and* same direction. Two vectors with the same magnitude but different directions are different vectors, as $\mathbf{A}$ and $\mathbf{B}$ are in Figure 3.1.

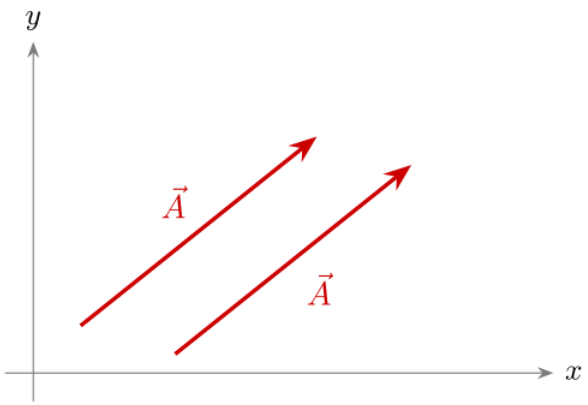


Figure 3.2. Two arrows of equal magnitude *and* direction represent the same vector, even when drawn in different locations. Only magnitude and direction matter, not the starting point, so both arrows are labeled $\mathbf{A}$.

•Figure description

Two arrows of identical length and identical direction drawn at different locations on x and y axes. Because only magnitude and direction define a vector, both arrows represent the same vector.

Representing vectors in two dimensions

A one-dimensional vector can be drawn on a number line. A scalar is represented by a single dot placed at the value of its magnitude; a one-dimensional vector is represented instead by an arrow, whose length gives its magnitude and whose orientation (left or right) gives its direction. A two-dimensional vector requires the xy-plane: we draw the vector as an arrow whose length is proportional to its magnitude and whose orientation indicates its direction. The arrow has a tail (its starting point) and a head (its tip).

Vectors are not pinned to any particular location. Two arrows of the same length pointing in the same direction represent the *same* vector, regardless of where on the page they are drawn (Figure 3.2). What identifies a vector is its magnitude and direction, not its position. This is why we may freely slide a vector around a diagram, as we will do constantly when adding vectors head-to-tail in §3.4.

This page titled 3.1: Scalars and Vectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.2: Components of a Vector

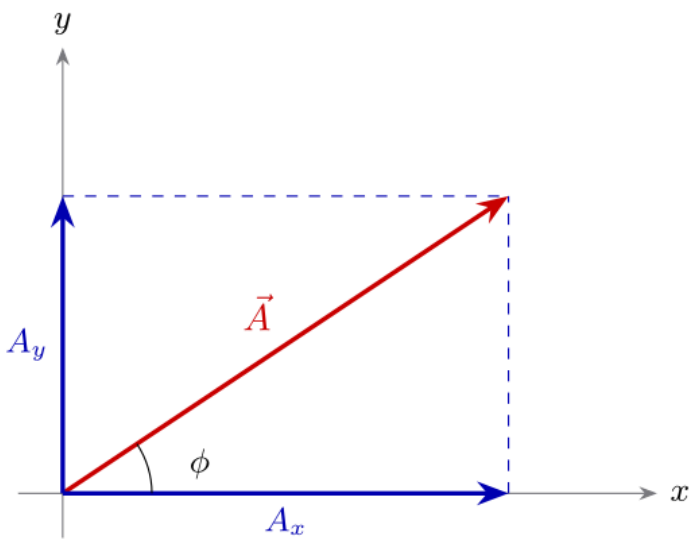


Figure 3.3. The components A_x and A_y are the signed projections of $\mathbf{A}$ onto the coordinate axes; the blue arrows are the corresponding component vectors $\mathbf{A}_x$ and $\mathbf{A}_y$. Dropping perpendiculars from the head of $\mathbf{A}$ forms a right triangle whose hypotenuse has length A and whose legs have lengths $|A_x|$ and $|A_y|$. Here ϕ is the acute angle between $\mathbf{A}$ and the $+x$-axis; the leg along that axis has length $A\cos\phi$ and the leg perpendicular to it has length $A\sin\phi$. The same construction works for an angle measured from *either* axis.

•Figure description

A vector A drawn from the origin on x and y axes, with dashed perpendicular lines dropped to each axis. The segments cut off along the axes are labeled A sub x and A sub y, and blue arrows along the axes show the corresponding component vectors.

A direction in the xy-plane is most economically described by giving an angle, but for calculation it is almost always easier to break the vector into pieces aligned with the coordinate axes. The **components** of $\mathbf{A}$ are the signed projections of $\mathbf{A}$ onto the axes: the "shadows" the vector casts on the x- and y-axes. We call them A_x and A_y. The word *signed* is essential: A_x is positive when the shadow points in the $+x$ direction and negative when it points in the $-x$ direction, and likewise for A_y. A component is a scalar, not a vector, but it carries a sign that records which way along the axis the vector leans.

Note: Component vectors versus scalar components

Two closely related objects share almost the same name, and confusing them is a common source of trouble.

The **scalar component** A_x is a signed number: for a vector pointing up and to the left, A_x might be $-3\ \mathrm{m}$. The **component vector** $\mathbf{A}_x = A_x\ \hat{\imath}$ is the arrow that this number describes: it has length $|A_x|$ and points along the x-axis. In Figure 3.3 the blue arrows are the component vectors; the labels A_x and A_y name their signed lengths.

The two are related by $\mathbf{A} = \mathbf{A}_x + \mathbf{A}_y$, which is a vector equation, and by $A_x = |\mathbf{A}_x|$ only when A_x happens to be positive. Whenever this book writes A_x without an arrow, a signed number is meant.

To find the components, drop a perpendicular from the head of $\mathbf{A}$ to each axis. This builds a right triangle whose hypotenuse has length A and whose two legs have lengths $|A_x|$ and $|A_y|$, as in Figure 3.3. Trigonometry then gives the *lengths* of the legs. The reliable way to do this (one that works no matter how the direction was specified) is:

1. **Sketch the vector** on the axes, roughly to scale, so you can see which way it points.
2. **Identify the acute angle the vector makes with one of the axes,** and call it ϕ. It does not matter which axis you measure from, and ϕ need not be measured from the $+x$-axis, but it must be the *acute* angle inside the right triangle you just drew, so that $0^\circ < \phi < 90^\circ$.
3. **Compute the leg lengths from that angle.** The leg lying *along* the axis you measured from is adjacent to ϕ, so its length is $A\cos\phi$; the leg *perpendicular* to that axis is opposite to ϕ, so its length is $A\sin\phi$.

4. **Attach the signs by looking at your sketch.** If a component points along $+x$ (or $+y$) it is positive; if it points along $-x$ (or $-y$) it is negative.

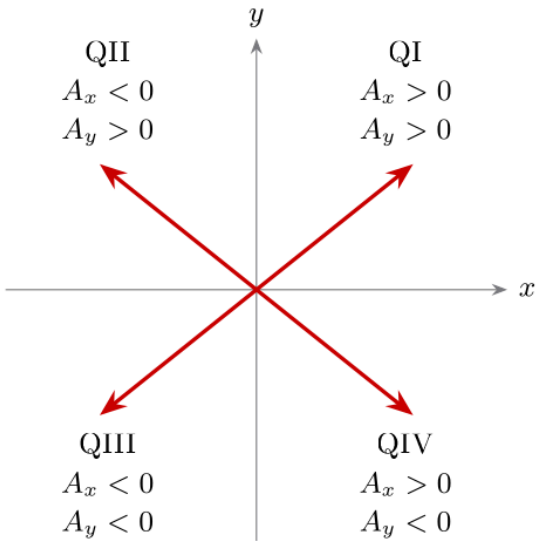


Figure 3.4. The signs of the components depend only on which quadrant the vector points into. The trigonometry of step 3 gives the *sizes* $|A_x|$ and $|A_y|$; this picture gives the signs. Sketching the vector first makes the signs obvious and costs a few seconds.

•Figure description

A set of x and y axes divided into four labeled quadrants. Quadrant one is marked A sub x positive and A sub y positive; quadrant two, A sub x negative and A sub y positive; quadrant three, both negative; quadrant four, A sub x positive and A sub y negative. The quadrant a vector points into fixes the signs of its components.

Step 4 is the one that makes the method general. The lengths from step 3 are never negative (they come from a right triangle), so the sign of each component is something *you* supply, by reading the picture, not something a formula hands you. Figure 3.4 shows the four possible sign patterns at a glance.

It is worth seeing how the familiar textbook formulas fit into this scheme. Let θ be the vector's **standard-position angle**: the angle measured counterclockwise from the $+x$-axis, running from 0° to 360°. For a vector in the first quadrant, θ coincides with the acute angle ϕ measured from the $+x$-axis, the x-leg is adjacent and the y-leg is opposite, and both components come out positive, so

$$A_x = A\cos\theta, \tag{3.1}$$
$$A_y = A\sin\theta. \tag{3.2}$$

In fact equations (3.1) and (3.2) remain correct in *every* quadrant, provided θ is the standard-position angle: the cosine and sine of an obtuse or reflex angle are already negative where they need to be, so the signs take care of themselves. That is a real convenience, and we will use it.

The catch is that a problem almost never hands you θ. It hands you something like "24° north of west" or "40° below the horizontal," and converting that into a standard-position angle is itself an error-prone step. So there are two honest routes:

- **Convert to a standard-position angle** θ, then apply (3.1) and (3.2) and let the trigonometric functions supply the signs.
- **Build the right triangle** on the acute angle you were actually given, compute the leg lengths, and supply the signs yourself from the sketch.

Both are correct, and both appear in the examples that follow. The second is usually faster and is far more forgiving when the axes are not horizontal and vertical. What is *not* safe is memorizing "x gets cosine, y gets sine" and applying it to whatever angle appears in the problem statement: that recipe is only valid for a standard-position angle, and it will betray you the first time an angle is measured from the y-axis. Either sketch the triangle, or make sure your θ really is measured counterclockwise from $+x$.

Going the other way (recovering the magnitude and direction from the components), the magnitude is always

Key Equation: Magnitude from components

$$A = \sqrt{A_x^2 + A_y^2} \quad \text{(two dimensions)}, \qquad A = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad \text{(three dimensions)}. \tag{3.3}$$

In two dimensions this is just the Pythagorean theorem applied to the right triangle of Figure 3.3; in three dimensions it is the Pythagorean theorem applied twice, first in the xy-plane and then in the vertical plane containing $\mathbf{A}$. Notice that squaring erases all the signs, so the magnitude comes out nonnegative no matter which quadrant (or octant) the vector points into, exactly as it must.

In two dimensions the direction is fixed by

$$\tan\theta = \frac{A_y}{A_x}, \tag{3.4}$$

but this one requires care. A calculator's inverse tangent returns an angle only in the range -90° to $+90^\circ$, so by itself it cannot distinguish a vector in the second quadrant from one in the fourth, or first from third. Always use the *signs* of A_x and A_y (equivalently, a glance at your sketch) to place the angle in the correct quadrant. In practice: compute $\tan^{-1}(A_y/A_x)$, look at your sketch, and if the calculator's answer points the wrong way, add 180°.

This projection idea is more general than it may first appear: it works for any pair of perpendicular axes, not only the horizontal and vertical ones. In Chapter 5, for example, we will tilt the axes to lie along and perpendicular to a ramp and resolve gravity into components in exactly the same way; build the right triangle on the angle you are given, then read off the signs.

Example 3.1: Resolving a force into components

A force of magnitude $50\ \text{N}$ acts at 20° above the horizontal. Find its x and y components.

Solution

Sketch the force: it points up and to the right, making a 20° angle with the horizontal (x) axis. Drop perpendiculars to the axes to form the right triangle. The leg along x is adjacent to the 20° angle and the leg along y is opposite it, so the leg lengths are

$$F\cos 20^\circ = (50)\cos 20^\circ \approx 47.0\,\text{N}, \qquad F\sin 20^\circ = (50)\sin 20^\circ \approx 17.1\,\text{N}. \tag{3.2.1}$$

From the sketch the force points in $+x$ and $+y$, so both components are positive:

$$F_x \approx +47.0\,\text{N}, \qquad F_y \approx +17.1\,\text{N}. \tag{3.2.2}$$

Example 3.2: Components from a compass direction

A car travels at $14\ \text{m/s}$ in a direction 24° north of west. Find the eastward and northward components of its velocity, taking east as $+x$ and north as $+y$.

Solution

"North of west" means: start pointing due west, then swing 24° toward north. Sketching this, the velocity points up and to the left (into the second quadrant) and makes a 24° angle with due west, that is, with the negative half of the x-axis. Build the right triangle on that 24° angle. The leg along the east–west (x) axis is adjacent, and the leg along the north–south (y) axis is opposite, so the leg lengths are

$$(14)\cos 24^\circ \approx 12.8\,\text{m/s}, \qquad (14)\sin 24^\circ \approx 5.7\,\text{m/s}. \tag{3.2.3}$$

Now read the signs from the sketch: the velocity points west ($-x$) and north ($+y$), so

$$v_x \approx -12.8\,\text{m/s}, \qquad v_y \approx +5.7\,\text{m/s}. \tag{3.2.4}$$

Notice that we used the 24° we were actually given and never had to convert the heading into an angle measured from the $+x$-axis. (Had we insisted on the standard-position angle $180^\circ - 24^\circ = 156^\circ$, then $14\cos 156^\circ$ and $14\sin 156^\circ$ would return the same two components, signs included, but the projection method gets there faster and with far less room for a sign error.)

Checkpoint 3.1

A vector $\mathbf{B}$ has $B_x = -3$ and $B_y = -4$. Without computing an inverse tangent, in which quadrant of the xy-plane does $\mathbf{B}$ lie? What is its magnitude? If you did feed $\tan\theta = B_y/B_x = (-4)/(-3)$ into a calculator's inverse tangent, why would the angle it returns be misleading?

Answer

Answers to Checkpoints

Checkpoint 3.1. Both components are negative, so $\mathbf{B}$ points to the left and down: it lies in the *third* quadrant. Its magnitude is $B = \sqrt{(-3)^2 + (-4)^2} = \sqrt{9 + 16} = 5$. The two minus signs cancel in the ratio, $(-4)/(-3) = +4/3$, so the calculator's inverse tangent returns 53.1°, a first-quadrant angle. That is the direction of a vector with components $(+3, +4)$, pointing exactly opposite to $\mathbf{B}$. The true direction is $53.1^\circ + 180^\circ = 233.1^\circ$ from the $+x$-axis, or 53.1° below the $-x$-axis. A sketch catches the error at once.

This page titled 3.2: Components of a Vector is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.3: Unit Vectors

A **unit vector** is a vector whose magnitude is exactly 1 (dimensionless). Its only job is to indicate a direction. Unit vectors are typically written with a "hat" rather than an arrow: $\hat{A}$ is read "A-hat."

The unit vector in the direction of a given vector

Given any nonzero vector $\mathbf{A}$, the unit vector pointing in the same direction is obtained by dividing $\mathbf{A}$ by its own magnitude:

$$\hat{A} = \frac{\mathbf{A}}{|\mathbf{A}|} = \frac{\mathbf{A}}{A}. \tag{3.5}$$

The division of a vector by a scalar (covered in detail in §3.5) just rescales the vector's length without changing its direction, so $\hat{A}$ points the same way as $\mathbf{A}$ but has length 1. Any vector can be written as the product of its magnitude and its direction:

$$\mathbf{A} = A\,\hat{A}. \tag{3.6}$$

Example 3.3: Unit vector from components

Given $\mathbf{A} = \langle -4\ \text{m}, 3\ \text{m}\rangle$, find the unit vector $\hat{A}$.

Solution

Step 1: Find the magnitude. By equation (3.3),

$$A = \sqrt{(-4)^2 + 3^2}\,\text{m} = \sqrt{16+9}\,\text{m} = \sqrt{25}\,\text{m} = 5\,\text{m}. \tag{3.3.1}$$

Keep the parentheses around -4 when you square it: $(-4)^2 = (-4)(-4) = +16$. Many calculators read -4^2 without parentheses as $-(4^2) = -16$, which would give the square root of a negative number here.

Step 2: Divide by the magnitude. Dividing a vector by a scalar divides *each* component by that scalar (equation (3.11) with $c = 1/A$):

$$\hat{A} = \frac{\mathbf{A}}{A} = \left\langle \frac{-4\,\text{m}}{5\,\text{m}},\ \frac{3\,\text{m}}{5\,\text{m}} \right\rangle = \langle -0.8,\ 0.6\rangle. \tag{3.3.2}$$

Every component shrank by the same factor of 5, so the direction is unchanged and the length went from 5 to 1. The meters cancel in each fraction, so $\hat{A}$ is dimensionless.

Check. $\sqrt{(-0.8)^2 + (0.6)^2} = \sqrt{0.64 + 0.36} = \sqrt{1.00} = 1$, so $\hat{A}$ has magnitude 1 as required.

Example 3.4: Unit vector from a compass direction

A force $\mathbf{C}$ has magnitude $450\ \text{N}$ and points 27° north of west. Find $\hat{C}$.

Solution

Take east as $+x$ and north as $+y$. This time we will convert to a standard-position angle: measuring counterclockwise from $+x$, the direction 27° north of west is at $\theta = 180^\circ - 27^\circ = 153^\circ$. The components of $\mathbf{C}$ are then

$$C_x = (450)\cos 153^\circ \approx -401\,\text{N}, \qquad C_y = (450)\sin 153^\circ \approx +204\,\text{N}, \tag{3.3.3}$$

and the signs come out right automatically, as promised. Then

$$\hat{C} = \frac{\mathbf{C}}{C} = \frac{\langle -401, 204\rangle\ \mathrm{N}}{450\ \mathrm{N}} \approx \langle -0.891,\ 0.454\rangle. \tag{3.3.4}$$

(Carry the unrounded components into this division; rounding to $204\ \mathrm{N}$ first would give 0.453 instead of 0.454. As a rule, round only at the very end of a calculation.) Again, the units of newtons cancel: a unit vector is dimensionless.

As a check, notice that $\hat{C} = \langle \cos 153^\circ, \sin 153^\circ \rangle$. That is no accident: dividing by the magnitude strips away the "how much" and leaves only the "which way," and for a two-dimensional vector the direction is completely captured by the cosine and sine of its standard-position angle.

The standard basis vectors

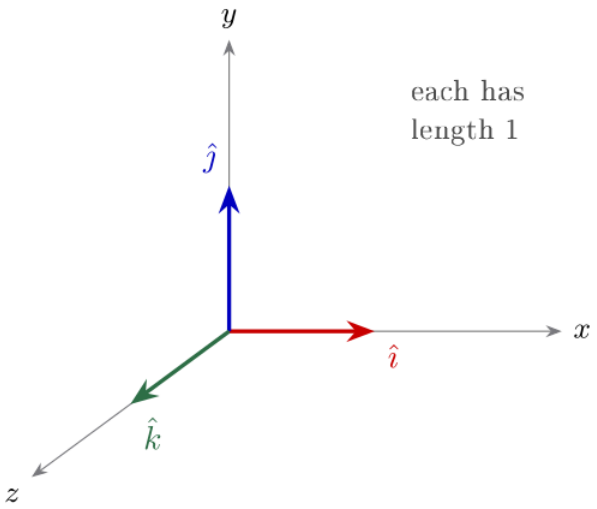


Figure 3.5. The standard basis vectors $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ point along the $+x$, $+y$, and $+z$ axes and each has magnitude 1. The axes form a *right-handed* set: curling the fingers of your right hand from $+x$ toward $+y$ makes your thumb point along $+z$. Every convention in this book, including the right-hand rule for cross products in §3.5, depends on the axes being labeled this way.

•Figure description

Three mutually perpendicular axes x, y and z drawn in perspective, each carrying a short unit arrow of length one: i-hat along positive x, j-hat along positive y, and k-hat along positive z. The axes form a right-handed set.

Particularly important are the unit vectors that point along the coordinate axes. We define

$$\hat{\imath} = \langle 1, 0, 0\rangle, \qquad \hat{\jmath} = \langle 0, 1, 0\rangle, \qquad \hat{k} = \langle 0, 0, 1\rangle, \tag{3.7}$$

called the **standard basis vectors** of three-dimensional space (Figure 3.5). In two dimensions, only $\hat{\imath}$ and $\hat{\jmath}$ are needed. These three mutually perpendicular unit vectors give us a third, very common way of writing vectors:

$$\mathbf{A} = A_x\,\hat{\imath} + A_y\,\hat{\jmath} + A_z\,\hat{k}. \tag{3.8}$$

So for example $\mathbf{A} = \langle 4, -3\rangle$ may equivalently be written as $\mathbf{A} = 4\hat{\imath} - 3\hat{\jmath}$. The three notations (magnitude and angle, ordered tuple of components, and basis-vector expansion) are completely interchangeable; we choose whichever is most convenient for the problem at hand.

Example 3.5: Three notations for one vector

A displacement vector is $\mathbf{C} = \langle 3\ \mathrm{m}, -4\ \mathrm{m}\rangle$. Write it in basis-vector form and in magnitude–angle form.

Solution

The basis-vector form just reads the components off as coefficients of $\hat{\imath}$ and $\hat{\jmath}$ (equation (3.8)):

$$\mathbf{C} = (3\hat{\imath} - 4\hat{\jmath})\,\mathrm{m}. \tag{3.3.5}$$

The minus sign belongs to the y-component: $+3\hat{\imath} + (-4)\hat{\jmath}$ is written more compactly as $3\hat{\imath} - 4\hat{\jmath}$. Writing the unit once, outside the parentheses, is cleaner than attaching "m" to each term.

For the magnitude–angle form, use equation (3.3), again keeping the negative component inside parentheses when it is squared:

$$C = \sqrt{3^2 + (-4)^2}\,\mathrm{m} = \sqrt{9 + 16}\,\mathrm{m} = \sqrt{25}\,\mathrm{m} = 5\,\mathrm{m}. \tag{3.3.6}$$

For the direction, equation (3.4) gives $\tan\theta = C_y/C_x = (-4)/3$. Before trusting a calculator, decide the quadrant from the signs: $C_x > 0$ and $C_y < 0$, so the vector points right and down, into the fourth quadrant. A calculator gives $\tan^{-1}(-4/3) \approx -53.1^\circ$, and a negative angle means clockwise from $+x$, which is indeed a fourth-quadrant direction, so no correction is needed:

$$\mathbf{C} = 5\,\mathrm{m}\ \text{at}\ 53.1^\circ\ \text{below the}\ +x\text{-axis}. \tag{3.3.7}$$

All three descriptions ($\langle 3, -4\rangle$ m, $(3\hat{\imath} - 4\hat{\jmath})$ m, and "5 m at -53.1°") refer to the same arrow. (The standard-position angle would be $360^\circ - 53.1^\circ = 306.9^\circ$, which names the same direction.)

Example 3.6: A three-dimensional vector

Write $\mathbf{D} = \langle 1\ \mathrm{m}, -2\ \mathrm{m}, 3\ \mathrm{m}\rangle$ in basis-vector form and find its magnitude.

Solution

$$\mathbf{D} = (\hat{\imath} - 2\hat{\jmath} + 3\hat{k})\,\mathrm{m}, \tag{3.3.8}$$

and by the three-dimensional form of equation (3.3),

$$D = \sqrt{1^2 + (-2)^2 + 3^2}\,\mathrm{m} = \sqrt{14}\,\mathrm{m} \approx 3.74\,\mathrm{m}. \tag{3.3.9}$$

Note that we do not attempt a magnitude–angle form here. In three dimensions a single angle is not enough to specify a direction, which is exactly why the unit vector $\hat{D} = \mathbf{D}/D$ is the tool of choice.

Checkpoint 3.2

Is $\hat{\imath} + \hat{\jmath}$ a unit vector? Justify your answer with a calculation.

Answer

Answers to Checkpoints

Checkpoint 3.2. No. Its magnitude is $|\hat{\imath} + \hat{\jmath}| = \sqrt{1^2 + 1^2} = \sqrt{2} \approx 1.41$, not 1. Adding two unit vectors does not, in general, give a unit vector. The unit vector pointing in the same direction is $(\hat{\imath} + \hat{\jmath})/\sqrt{2}$.

This page titled 3.3: Unit Vectors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.4: Vector Addition

There are several ways to add two vectors, but only one is genuinely useful for calculation.

Three methods, one workhorse

Graphically. One can literally draw both vectors to scale on a piece of paper, place the tail of the second vector at the head of the first, and measure the resulting "head-to-tail" sum with a ruler and protractor. This works in principle, but the precision is dictated by your draftsmanship, and rarely better than a couple of percent. It remains valuable as a *check*: a quick sketch tells you roughly where the answer should land, which is often enough to catch a sign error in the arithmetic.

Geometrically. Knowing the magnitudes of the two vectors and the angle between them, one can find the sum using the law of cosines and the law of sines. This is exact, but tedious, and the geometry becomes painful for more than two vectors.

By components. If we resolve each vector into its x, y, and z components, then *the components add component-by-component*. This reduces vector addition to ordinary scalar arithmetic and generalizes immediately to any number of vectors and any number of dimensions. This is the method we use.

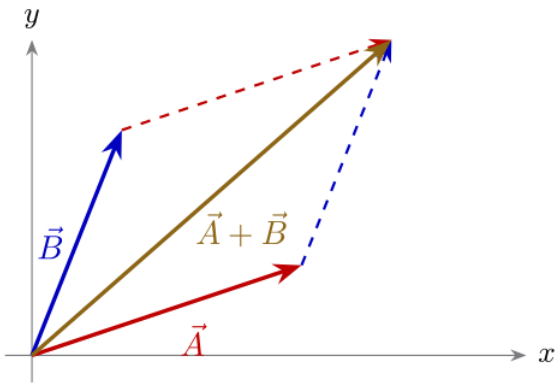


Figure 3.6. Adding two vectors using the parallelogram (or head-to-tail) construction. Sliding $\mathbf{B}$ so its tail sits at the head of $\mathbf{A}$, or sliding $\mathbf{A}$ so its tail sits at the head of $\mathbf{B}$, leads to the same corner: the two dashed routes give the same sum, which is why $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$. The sum is the diagonal drawn from the common tail; the *other* diagonal of the parallelogram, joining the two heads, is $\mathbf{A}-\mathbf{B}$ (see Figure 3.7). The actual calculation is done by adding components, not by drawing.

•Figure description

Two vectors A and B drawn from a common origin, completed into a parallelogram by dashed lines. The diagonal from the origin is the sum A plus B, matching the head-to-tail construction in which B is slid so its tail meets the head of A.

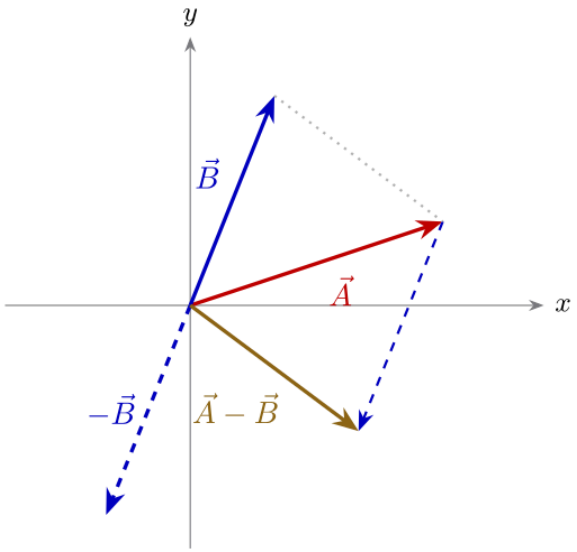


Figure 3.7. Subtracting vectors. Reverse $\mathbf{B}$ to get $-\mathbf{B}$, then add head-to-tail. Equivalently, $\mathbf{A}-\mathbf{B}$ is the arrow running from the head of $\mathbf{B}$ to the head of $\mathbf{A}$ (dotted), which is the second diagonal of the parallelogram in Figure 3.6. Note that here $|\mathbf{A}-\mathbf{B}|$ is *larger* than $|\mathbf{A}|$: subtraction does not necessarily shrink a vector.

•Figure description

Vector subtraction shown two ways. Vector B is reversed to give minus B and added head-to-tail to A, and equivalently an arrow is drawn from the head of B to the head of A. Both give A minus B.

If $\mathbf{A}=\langle A_x, A_y, A_z\rangle$ and $\mathbf{B}=\langle B_x, B_y, B_z\rangle$, then

Key Equation: Vector addition by components

$$\mathbf{A}+\mathbf{B}=\langle A_x+B_x,\ A_y+B_y,\ A_z+B_z\rangle. \tag{3.9}$$

Vector subtraction works the same way: $\mathbf{A}-\mathbf{B}=\mathbf{A}+(-\mathbf{B})$, where $-\mathbf{B}$ has the same magnitude as $\mathbf{B}$ but the opposite direction (every component negated). So

$$\mathbf{A}-\mathbf{B}=\langle A_x-B_x,\ A_y-B_y,\ A_z-B_z\rangle. \tag{3.10}$$

Subtraction is worth a picture of its own, because students often expect $\mathbf{A}-\mathbf{B}$ to be shorter than $\mathbf{A}$, and it need not be. Figure 3.7 shows the construction.

Example 3.7: A plane's net displacement

A plane flies $45\ \text{mi}$ at 30° east of north, then turns and flies $55\ \text{mi}$ at 40° south of east. How far is the plane from where it started?

Solution

Take east as $+x$ and north as $+y$, and sketch both legs before computing anything.

First leg. "30° east of north" means: start pointing north, then swing 30° toward east. The leg points up and to the right, into the first quadrant, and makes a 30° angle with the $+y$-axis. Build the right triangle on that 30° angle. Because we measured from the y-axis, the y-leg is now the adjacent one (cosine) and the x-leg is opposite (sine):

$$|A_y| = 45\cos 30^\circ \approx 39.0\,\text{mi}, \qquad |A_x| = 45\sin 30^\circ = 22.5\,\text{mi}. \tag{3.4.1}$$

($\sin 30^\circ$ is exactly $\frac{1}{2}$, so A_x is exactly half of 45.) Both point in positive directions, so

$$\mathbf{A}=\langle 22.5,\ 39.0\rangle\ \text{mi}. \tag{3.4.2}$$

This is exactly the situation warned about earlier: the angle was given from the y-axis, so cosine landed on y and sine on x. Blindly writing $A_x = 45\cos 30^\circ$ would have been wrong.

Second leg. "40° south of east" points down and to the right, into the fourth quadrant, making a 40° angle with the $+x$-axis. Here the x-leg is adjacent:

$$|B_x| = 55\cos 40^\circ \approx 42.1\,\text{mi}, \qquad |B_y| = 55\sin 40^\circ \approx 35.4\,\text{mi}. \tag{3.4.3}$$

From the sketch, B_x is eastward $(+)$ and B_y is southward $(-)$:

$$\mathbf{B} = \langle 42.1,\ -35.4\rangle\,\text{mi}. \tag{3.4.4}$$

Add. By equation (3.9), add the x-components to each other and the y-components to each other, never an x to a y:

$$\mathbf{R} = \mathbf{A} + \mathbf{B} = \langle 22.5 + 42.1,\ 39.0 + (-35.4)\rangle\,\text{mi} = \langle 64.6,\ 3.6\rangle\,\text{mi}. \tag{3.4.5}$$

Magnitude and direction. The components are the legs of a right triangle whose hypotenuse is R:

$$R = \sqrt{(64.6)^2 + (3.6)^2}\,\text{mi} = \sqrt{4173 + 13}\,\text{mi} \approx 64.7\,\text{mi}. \tag{3.4.6}$$

The plane ends up about $64.7\ \text{mi}$ from its starting point, very nearly due east of it: $\theta = \tan^{-1}(3.6/64.6) \approx 3.2^\circ$ north of east. Both components are positive, so the vector is in the first quadrant and the calculator's answer needs no correction.

Check. Two legs of 45 and $55\ \text{mi}$ can put the plane no farther than $45 + 55 = 100\ \text{mi}$ from the start (if they point the same way) and no closer than $55 - 45 = 10\ \text{mi}$ (if they point opposite ways). Our $64.7\ \text{mi}$ lies in that range, and the sketch (one leg angled north, the other angled south by a similar amount) agrees that the result should point nearly east.

Notice that the small northward component, $3.6\ \text{mi}$, is the difference of two numbers of size roughly $37\ \text{mi}$. Differences like this lose precision quickly, so it is worth carrying extra digits until the end.

Example 3.8: Net force on a shopping cart

A man pulls a shopping cart directly to the right with a force of $500\ \text{N}$. A woman pulls the same cart with a force of $700\ \text{N}$ to the left at an angle of 15° above the horizontal. Find the combined force exerted by the two people.

Solution

Take rightward as $+x$ and upward as $+y$. This time we will convert to standard-position angles and let the trigonometric functions handle the signs. The man's force, directed in the $+x$ direction, is

$$\mathbf{F}_\text{M} = \langle 500,\ 0\rangle\,\text{N}. \tag{3.4.7}$$

The woman's force is to the left $(-x)$ at 15° above the horizontal $(+y)$. Measured counterclockwise from $+x$, straight left is 180°; tipping 15° back up toward $+y$ brings it to $180^\circ - 15^\circ = 165^\circ$:

$$\mathbf{F}_\text{W} = \langle 700\cos 165^\circ,\ 700\sin 165^\circ\rangle\,\text{N} \approx \langle -676,\ 181\rangle\,\text{N}. \tag{3.4.8}$$

The cosine of 165° is negative and its sine is positive, exactly the signs a second-quadrant vector needs. The combined force is the component-by-component sum:

$$\mathbf{F}_\text{M} + \mathbf{F}_\text{W} \approx \langle 500 + (-676),\ 0 + 181\rangle\,\text{N} = \langle -176,\ 181\rangle\,\text{N}. \tag{3.4.9}$$

Its magnitude is

$$\sqrt{(-176)^2 + (181)^2}\,\text{N} \approx 253\,\text{N}. \tag{3.4.10}$$

(Squaring makes both terms positive, so the sign of F_x does not affect the magnitude. The unrounded components, -176.2 and 181.2, give $252.7\ \mathrm{N}$; the rounded ones would give $252.5\ \mathrm{N}$: another reason to round only at the end.) For the direction, be careful. Feeding $181/(-176)$ into a calculator's inverse tangent returns -45.8°, which points down and to the right. But $F_x < 0$ and $F_y > 0$, so the vector lies in the *second* quadrant, up and to the left. Adding 180° puts it there:

$$\theta = -45.8^\circ + 180^\circ = 134.2^\circ. \tag{3.4.11}$$

Adding 180° works because $\tan\theta$ repeats every 180°: the angles -45.8° and 134.2° have the same tangent but point in opposite directions, and only the sketch can tell them apart. This is precisely the trap described after equation (3.4). Had we simply written down the calculator's -45.8°, our answer would have pointed in exactly the wrong direction. The woman is winning the tug-of-war by a substantial margin.

Check. The man's $500\ \mathrm{N}$ nearly cancels the woman's $676\ \mathrm{N}$ horizontal pull, leaving about $176\ \mathrm{N}$ to the left, while nothing opposes her $181\ \mathrm{N}$ upward pull. Two roughly equal perpendicular components should give a direction close to 45° from the horizontal, here $180^\circ - 134.2^\circ = 45.8^\circ$ above the leftward direction.

Note: "Net force" means *all* the forces

The $253\ \mathrm{N}$ above is the resultant of the two *pulls* only. It is not the net force on the cart, because the cart also experiences its weight, the normal force from the floor, and friction. Starting in Chapter 5, "net force" will always mean the vector sum of *every* force acting on an object, and leaving one out is one of the most common ways to get a wrong answer. Here we are practicing the vector addition itself, not yet doing dynamics.

This page titled 3.4: Vector Addition is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.5: Vector Multiplication

There are three kinds of "multiplication" involving vectors: a vector times a scalar, a vector times another vector to produce a scalar (the dot product), and a vector times another vector to produce a third vector (the cross product). All three appear throughout physics. There is no operation that divides one vector by another.

Scalar multiplication

Multiplying a vector $\mathbf{A}$ by a scalar c scales the vector's magnitude by $|c|$ and either preserves its direction (if $c > 0$) or reverses it (if $c < 0$). Component-wise:

$$c\,\mathbf{A} = \langle cA_x,\ cA_y,\ cA_z \rangle. \tag{3.11}$$

The product $c\mathbf{A}$ is itself a vector. Division by a scalar is the same as multiplication by its reciprocal; this is what we did in equation (3.5) when we built $\hat{A} = \mathbf{A}/A$.

Example 3.9: Scaling a vector

Given $\mathbf{A} = \langle 4, -3 \rangle$, find $4\mathbf{A}$ and $-\frac{1}{2}\mathbf{A}$.

Solution

$$4\mathbf{A} = \langle 4(4),\ 4(-3) \rangle = \langle 16,\ -12 \rangle. \tag{3.5.1}$$

$$-\tfrac{1}{2}\mathbf{A} = \langle -\tfrac{1}{2}(4),\ -\tfrac{1}{2}(-3) \rangle = \langle -2,\ 1.5 \rangle. \tag{3.5.2}$$

The first quadruples the magnitude in the same direction; the second halves the magnitude and reverses the direction.

The dot product

The **dot product** (also called the *scalar product* or *inner product*) takes two vectors and returns a scalar. It answers the question "how much of one vector points along the other?", which is exactly the question that arises when a force does work on an object moving in some direction. You will meet this product again in Chapter 7, where work is defined as $W = \mathbf{F} \cdot \mathbf{d}$.

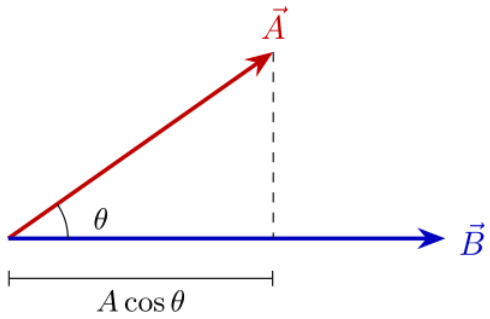


Figure 3.8. Geometric interpretation of the dot product. $A\cos\theta$ is the component of $\mathbf{A}$ along the direction of $\mathbf{B}$; multiplying it by B gives $\mathbf{A} \cdot \mathbf{B}$.
•Figure description
Two vectors A and B from a common origin with angle theta between them. A dashed perpendicular drops from the head of A onto the line of B, marking the segment A cosine theta, the component of A along B.

Geometrically, the dot product equals the product of the two magnitudes times the cosine of the angle between them:

Key Equation: Dot product

$$\mathbf{A}\cdot\mathbf{B} = AB\cos\theta = A_xB_x + A_yB_y + A_zB_z, \tag{3.12}$$

where θ is the angle between $\mathbf{A}$ and $\mathbf{B}$. The first equality is the geometric definition; the second is the equivalent component formula. Both always give the same answer. Geometrically, $A\cos\theta$ is the length of the projection of $\mathbf{A}$ onto the line of $\mathbf{B}$ (Figure 3.8); the dot product can therefore be read as "the projection of one vector onto the other, times the other's magnitude."

Several useful properties follow directly from equation (3.12):

- The dot product is *commutative*: $\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}$.
- It is *distributive* over addition: $\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}$. Scalars pull out freely: $(c\mathbf{A})\cdot\mathbf{B} = c(\mathbf{A}\cdot\mathbf{B})$.
- If $\mathbf{A}$ and $\mathbf{B}$ are perpendicular, $\theta = 90^\circ$, so $\mathbf{A}\cdot\mathbf{B} = 0$. This is often the easiest way to test for perpendicularity.
- The sign of the dot product tells you whether the angle is acute ($\mathbf{A}\cdot\mathbf{B} > 0$) or obtuse ($\mathbf{A}\cdot\mathbf{B} < 0$).
- $\mathbf{A}\cdot\mathbf{A} = A^2$, since the angle is 0° and $\cos 0^\circ = 1$. The dot product of a vector with itself is the square of its magnitude.
- For the basis vectors: $\hat{\imath}\cdot\hat{\imath} = \hat{\jmath}\cdot\hat{\jmath} = \hat{k}\cdot\hat{k} = 1$, while $\hat{\imath}\cdot\hat{\jmath} = \hat{\jmath}\cdot\hat{k} = \hat{k}\cdot\hat{\imath} = 0$.

The angle between two vectors

Equation (3.12) contains two expressions for the same number, and setting them equal and solving for $\cos\theta$ gives one of the most useful formulas in the chapter:

Key Equation: Angle between two vectors

$$\cos\theta = \frac{\mathbf{A}\cdot\mathbf{B}}{AB} = \frac{A_xB_x + A_yB_y + A_zB_z}{\sqrt{A_x^2 + A_y^2 + A_z^2}\,\sqrt{B_x^2 + B_y^2 + B_z^2}}. \tag{3.13}$$

This works in any number of dimensions and always returns θ between 0° and 180°, which is precisely the range we want, since "the angle between two vectors" is never taken to be more than a straight angle. Unlike the inverse tangent of §3.2, the inverse cosine here needs no quadrant correction.

Example 3.10: Angle between two vectors

Find the angle between $\mathbf{A} = \langle 2, 1, -2\rangle$ and $\mathbf{B} = \langle 0, 3, 4\rangle$.

Solution

Step 1: The dot product. Multiply matching components (x with x, y with y, z with z) and add the three products:

$$\mathbf{A}\cdot\mathbf{B} = (2)(0) + (1)(3) + (-2)(4) = 0 + 3 - 8 = -5. \tag{3.5.3}$$

Step 2: The magnitudes. Square each component, add, and take the square root:

$$\begin{aligned} A &= \sqrt{2^2 + 1^2 + (-2)^2} = \sqrt{4 + 1 + 4} = \sqrt{9} = 3, \\ B &= \sqrt{0^2 + 3^2 + 4^2} = \sqrt{0 + 9 + 16} = \sqrt{25} = 5. \end{aligned} \tag{3.5.4}$$

Step 3: Solve for the angle. By equation (3.13),

$$\cos\theta = \frac{\mathbf{A}\cdot\mathbf{B}}{AB} = \frac{-5}{(3)(5)} = -\frac{5}{15} = -\frac{1}{3}. \tag{3.5.5}$$

To undo the cosine, take the inverse cosine of both sides:

$$\theta = \cos^{-1}\!\left(-\tfrac{1}{3}\right) \approx 109.5^\circ. \tag{3.5.6}$$

Unlike the inverse tangent, the calculator's inverse cosine already returns an angle between 0° and 180°, so there is no quadrant to fix.

Check. The negative dot product told us in advance that $\cos\theta < 0$, so the angle had to be obtuse (between 90° and 180°), and it is.

Projection of one vector onto another

Figure 3.8 shows that $A\cos\theta$ is the length of the shadow $\mathbf{A}$ casts on the line of $\mathbf{B}$. Dividing equation (3.12) by B isolates that shadow, and writing $\hat{B} = \mathbf{B}/B$ puts it in its most useful form. The **scalar projection** of $\mathbf{A}$ onto $\mathbf{B}$ is

$$A_{\parallel} = A\cos\theta = \mathbf{A}\cdot\hat{B}, \tag{3.14}$$

a signed number: it is negative when $\mathbf{A}$ leans backward against $\mathbf{B}$. Attaching the direction $\hat{B}$ to it gives the **vector projection**, the piece of $\mathbf{A}$ that lies along $\mathbf{B}$:

Key Equation: Vector projection

$$\mathbf{A}_{\parallel} = (\mathbf{A}\cdot\hat{B})\,\hat{B}. \tag{3.15}$$

Whatever is left over is perpendicular to $\mathbf{B}$:

$$\mathbf{A}_{\perp} = \mathbf{A} - \mathbf{A}_{\parallel}. \tag{3.16}$$

These two pieces are perpendicular to each other and add back to $\mathbf{A}$, so by the Pythagorean theorem $A_{\parallel}^2 + A_{\perp}^2 = A^2$: a convenient way to check your work.

This is the same idea as resolving a vector into components in §3.2, generalized to an axis pointing in any direction at all rather than along x or y. It is how we will split a weight into pieces along and perpendicular to a ramp in Chapter 5.

Example 3.11: Splitting a vector along and across a direction

Find the components of $\mathbf{A} = \langle 5, 2, 0\rangle$ parallel and perpendicular to $\mathbf{B} = \langle 0, 3, 4\rangle$.

Solution

Step 1: Build the unit vector along $\mathbf{B}$. Only the *direction* of $\mathbf{B}$ matters here, so strip away its length. Since $B = \sqrt{0+9+16} = 5$, divide each component by 5:

$$\hat{B} = \frac{\mathbf{B}}{B} = \left\langle \frac{0}{5},\ \frac{3}{5},\ \frac{4}{5}\right\rangle = \langle 0,\ 0.6,\ 0.8\rangle. \tag{3.5.7}$$

Step 2: The scalar projection. By equation (3.14), dot $\mathbf{A}$ into $\hat{B}$:

$$A_{\parallel} = \mathbf{A}\cdot\hat{B} = (5)(0) + (2)(0.6) + (0)(0.8) = 0 + 1.2 + 0 = 1.2. \tag{3.5.8}$$

This is the signed length of the piece of $\mathbf{A}$ that lies along $\mathbf{B}$. It is positive, so that piece points *with* $\mathbf{B}$, not against it.

Step 3: The vector projection. Attach the direction by multiplying the unit vector by this length (scalar multiplication, equation (3.11)):

$$\mathbf{A}_{\parallel} = 1.2\,\langle 0,\ 0.6,\ 0.8\rangle = \langle 1.2(0),\ 1.2(0.6),\ 1.2(0.8)\rangle = \langle 0,\ 0.72,\ 0.96\rangle. \tag{3.5.9}$$

Step 4: The perpendicular piece. Whatever is left over is perpendicular. Subtract component by component, equation (3.16):

$$\mathbf{A}_{\perp} = \mathbf{A} - \mathbf{A}_{\parallel} = \langle 5-0,\ 2-0.72,\ 0-0.96\rangle = \langle 5,\ 1.28,\ -0.96\rangle. \tag{3.5.10}$$

Note that the z-component of $\mathbf{A}_\perp$ is negative even though $\mathbf{A}$ had no z-component at all: the parallel piece leans in the $+z$ direction, so the leftover must lean the other way to cancel it.

Check 1: perpendicular really means perpendicular. If $\mathbf{A}_\perp$ is perpendicular to $\mathbf{B}$, their dot product must be zero:

$$\mathbf{A}_\perp \cdot \mathbf{B} = (5)(0) + (1.28)(3) + (-0.96)(4) = 0 + 3.84 - 3.84 = 0. \checkmark \tag{3.5.11}$$

Check 2: the pieces rebuild $\mathbf{A}$. Because the two pieces are perpendicular, the Pythagorean theorem applies to their lengths:

$$\begin{aligned} A_\parallel^2 + A_\perp^2 &= (1.2)^2 + \left[5^2 + (1.28)^2 + (-0.96)^2\right] \\ &= 1.44 + (25 + 1.638 + 0.922) = 1.44 + 27.56 = 29, \end{aligned} \tag{3.5.12}$$

and $A^2 = 5^2 + 2^2 + 0^2 = 29$. ✓

The cross product

The **cross product** (also called the *vector product*) takes two vectors and returns a third vector perpendicular to both of them. Where the dot product measures how much two vectors point *along* each other, the cross product measures how much they point *across* each other. It is the natural language for turning effects, and it returns in Chapter 10 as torque, $\tau = \mathbf{r} \times \mathbf{F}$, and in Chapter 11 as angular momentum.

Its magnitude is

Key Equation: Cross product magnitude

$$|\mathbf{A} \times \mathbf{B}| = AB\sin\theta, \tag{3.17}$$

where θ is the angle between the two vectors, taken between 0° and 180°, the same convention as for the dot product, and the one equation (3.13) automatically produces. On that range $\sin\theta$ is never negative, so the magnitude is never negative, as a magnitude must not be.

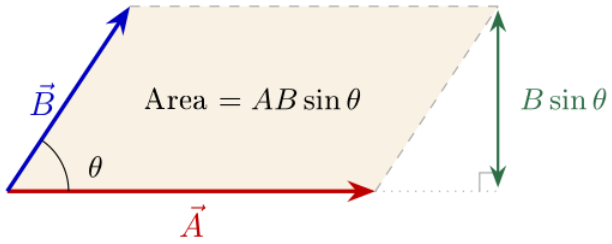


Figure 3.9. The magnitude of $\mathbf{A} \times \mathbf{B}$ equals the area of the shaded parallelogram: base A times height $B\sin\theta$. The cross product vector itself is perpendicular to the page and, for the ordering $\mathbf{A} \times \mathbf{B}$ shown here, points out of the page toward you. A triangle with sides $\mathbf{A}$ and $\mathbf{B}$ is half of this parallelogram, so its area is $\frac{1}{2}|\mathbf{A} \times \mathbf{B}|$.

•Figure description

Two vectors A and B from a common origin, with the parallelogram they span shaded. The base is labeled A and the perpendicular height B sine theta, so the shaded area equals A B sine theta, the magnitude of the cross product.

Geometrically, this magnitude equals the area of the parallelogram whose sides are $\mathbf{A}$ and $\mathbf{B}$ (Figure 3.9). This is just "base times height": taking $\mathbf{A}$ as the base of length A, the height of the parallelogram is $B\sin\theta$.

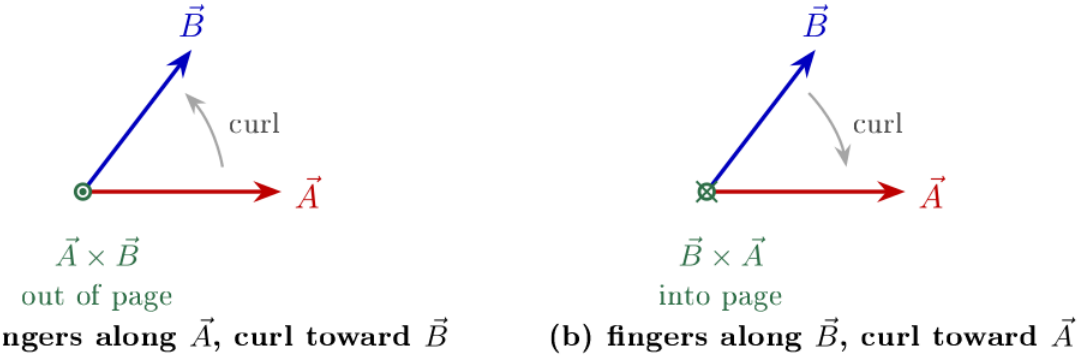


Figure 3.10. The right-hand rule. *Left:* point your right fingers along $\mathbf{A}$ and curl them toward $\mathbf{B}$; your thumb points out of the page, shown by the dot $\odot$. *Right:* reversing the order reverses the result; curling from $\mathbf{B}$ toward $\mathbf{A}$ sends your thumb into the page, shown by the cross $\otimes$. Same magnitude, opposite direction.

•Figure description

Two side-by-side diagrams of the right-hand rule. On the left, fingers curl from vector A toward vector B and the resulting cross product points out of the page, marked by a circled dot. On the right the order is reversed, and the result points into the page, marked by a circled cross.

The **direction** of $\mathbf{A} \times \mathbf{B}$ is perpendicular to the plane containing $\mathbf{A}$ and $\mathbf{B}$, and is determined by the **right-hand rule**: point the fingers of your right hand along $\mathbf{A}$, curl them toward $\mathbf{B}$ through the smaller angle, and your thumb points in the direction of $\mathbf{A} \times \mathbf{B}$ (Figure 3.10). Use your right hand: the left hand gives exactly the wrong answer every time.

Because the result is perpendicular to the page whenever $\mathbf{A}$ and $\mathbf{B}$ lie in it, we need a way to draw a vector pointing toward or away from the reader. The standard notation is an arrow seen from either end: a dot $\odot$ for an arrowhead coming toward you, and a cross $\otimes$ for the tail feathers going away from you.

Equivalently, $\mathbf{A} \times \mathbf{B}$ and $\mathbf{B} \times \mathbf{A}$ have the same magnitude but opposite directions:

$$\mathbf{B} \times \mathbf{A} = -(\mathbf{A} \times \mathbf{B}). \tag{3.18}$$

The cross product is therefore *not* commutative; the order matters.

In component form, the cross product is most easily computed as a 3×3 determinant:

Key Equation: Cross product by components

$$\mathbf{A} \times \mathbf{B} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} = (A_y B_z - A_z B_y)\,\hat{\imath} - (A_x B_z - A_z B_x)\,\hat{\jmath} + (A_x B_y - A_y B_x)\,\hat{k}. \tag{3.19}$$

If both vectors lie in the xy-plane ($A_z = B_z = 0$), the cross product has only a z-component, and equation (3.19) reduces to

$$\mathbf{A} \times \mathbf{B} = (A_x B_y - A_y B_x)\,\hat{k}. \tag{3.5.13}$$

This z-component is positive when $\mathbf{A} \times \mathbf{B}$ points out of the page (toward you) and negative when it points into the page.

Like the dot product, the cross product has a few key special cases worth memorizing:

- If $\mathbf{A}$ and $\mathbf{B}$ are parallel or antiparallel, $\sin\theta = 0$ and $\mathbf{A} \times \mathbf{B} = \mathbf{0}$. In particular, $\mathbf{A} \times \mathbf{A} = \mathbf{0}$. Note that this is the zero *vector*, not the scalar 0.
- The cross product is distributive, $\mathbf{A} \times (\mathbf{B} + \mathbf{C}) = \mathbf{A} \times \mathbf{B} + \mathbf{A} \times \mathbf{C}$, and scalars pull out: $(c\mathbf{A}) \times \mathbf{B} = c(\mathbf{A} \times \mathbf{B})$. But it is *not* associative, so parentheses matter.
- For the basis vectors: $\hat{\imath} \times \hat{\jmath} = \hat{k}$, $\hat{\jmath} \times \hat{k} = \hat{\imath}$, and $\hat{k} \times \hat{\imath} = \hat{\jmath}$ (and the reverse orderings give minus the result, by (3.18)). These cycle in the order $x \to y \to z \to x$.
- Because $\mathbf{A} \times \mathbf{B}$ is perpendicular to both inputs, $\mathbf{A} \cdot (\mathbf{A} \times \mathbf{B}) = 0$ and $\mathbf{B} \cdot (\mathbf{A} \times \mathbf{B}) = 0$ always. This is a fast check on any cross

product you compute.

Two geometric consequences are worth recording, since problems ask for them constantly. A triangle with sides $\mathbf{A}$ and $\mathbf{B}$ is exactly half of the parallelogram in Figure 3.9, so

$$\text{Area of triangle} = \tfrac{1}{2}\,|\mathbf{A} \times \mathbf{B}|. \tag{3.20}$$

And combining the two products gives the **scalar triple product** $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$, whose absolute value is the volume of the parallelepiped with edges $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. Note that the parentheses are not optional in spirit but they are in practice: $(\mathbf{A} \cdot \mathbf{B}) \times \mathbf{C}$ is meaningless, because you cannot take the cross product of a scalar with a vector, so the only sensible reading of $\mathbf{A} \cdot \mathbf{B} \times \mathbf{C}$ is the one with the cross product done first.

Example 3.12: Dot and cross products by components

A force $\mathbf{F} = \langle 3, 2, -2 \rangle$ N is applied at a point whose position relative to a pivot is $\mathbf{r} = \langle -2, 4, 5 \rangle$ m. Compute:

a. $\mathbf{r} \cdot \mathbf{F}$
b. $\mathbf{F} \cdot \mathbf{r}$
c. $3(\mathbf{r} \cdot \mathbf{F})$
d. $\mathbf{F} \times \mathbf{r}$
e. $\mathbf{r} \times \mathbf{F}$, which is the torque about the pivot

Solution

(a) Using the component form of equation (3.12), multiply matching components and add:

$$\mathbf{r} \cdot \mathbf{F} = (-2)(3) + (4)(2) + (5)(-2) = -6 + 8 - 10 = -8\,\text{N} \cdot \text{m}. \tag{3.5.14}$$

Each product is (meters)(newtons), so the result is in newton-meters.

(b) By commutativity of the dot product,

$$\mathbf{F} \cdot \mathbf{r} = \mathbf{r} \cdot \mathbf{F} = -8\,\text{N} \cdot \text{m}. \tag{3.5.15}$$

(Recomputing with the factors swapped, $(3)(-2) + (2)(4) + (-2)(5)$, gives the same three products in the same order, since ordinary multiplication doesn't care which number comes first.)

(c) A dot product is just a number, so multiplying it by 3 is ordinary arithmetic:

$$3(\mathbf{r} \cdot \mathbf{F}) = 3(-8)\,\text{N} \cdot \text{m} = -24\,\text{N} \cdot \text{m}. \tag{3.5.16}$$

(d) Set up the determinant of equation (3.19). The *first* vector in the product goes in the middle row and the *second* in the bottom row, so here $\mathbf{F}$ is in the middle and $\mathbf{r}$ is on the bottom:

$$\mathbf{F} \times \mathbf{r} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 3 & 2 & -2 \\ -2 & 4 & 5 \end{vmatrix}. \tag{3.5.17}$$

Each component comes from a 2×2 determinant. To find the one that multiplies a given unit vector, cover up that unit vector's row and column; the four numbers left over form the 2×2 determinant, which is evaluated as

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{3.5.18}$$

(the product down the main diagonal minus the product up the other diagonal). The middle term, for $\hat{\jmath}$, carries an extra minus sign. Covering the first, middle, and last columns in turn gives

$$\begin{aligned}
\hat{\imath}: &\quad \begin{vmatrix} 2 & -2 \\ 4 & 5 \end{vmatrix} = (2)(5) - (-2)(4) = 10 - (-8) = 10 + 8 = 18, \\
[4pt]\hat{\jmath}: &\quad -\begin{vmatrix} 3 & -2 \\ -2 & 5 \end{vmatrix} = -\big[(3)(5) - (-2)(-2)\big] = -(15 - 4) = -11, \\
[4pt]\hat{k}: &\quad \begin{vmatrix} 3 & 2 \\ -2 & 4 \end{vmatrix} = (3)(4) - (2)(-2) = 12 - (-4) = 12 + 4 = 16.
\end{aligned} \tag{3.5.19}$$

The negative numbers are where the errors hide. In the $\hat{\imath}$ term, $(-2)(4) = -8$, and subtracting -8 *adds* 8. In the $\hat{\jmath}$ term, $(-2)(-2) = +4$ because two negatives multiply to a positive, so we subtract 4. Assembling the three components,

$$\mathbf{F} \times \mathbf{r} = (18\,\hat{\imath} - 11\,\hat{\jmath} + 16\,\hat{k})\,\mathrm{N \cdot m} = \langle 18,\ -11,\ 16\rangle\,\mathrm{N \cdot m}. \tag{3.5.20}$$

(e) By anticommutativity, equation (3.18), reversing the order flips the sign of every component, so we get the torque for free:

$$\tau = \mathbf{r} \times \mathbf{F} = -(\mathbf{F} \times \mathbf{r}) = \langle -18,\ 11,\ -16\rangle\,\mathrm{N \cdot m}. \tag{3.5.21}$$

Two things to notice. First, the order matters enormously: parts (d) and (e) differ by a sign throughout, and only (e) is the torque. Second, both products come out in newton-meters, but they mean different things: the cross product is a torque, while the dot product would be a work if $\mathbf{r}$ were a displacement.

Check. A cross product is perpendicular to both of its inputs, so its dot product with either one must vanish:

$$\mathbf{r} \cdot (\mathbf{r} \times \mathbf{F}) = (-2)(-18) + (4)(11) + (5)(-16) = 36 + 44 - 80 = 0,\ \checkmark \tag{3.5.22}$$

$$\mathbf{F} \cdot (\mathbf{r} \times \mathbf{F}) = (3)(-18) + (2)(11) + (-2)(-16) = -54 + 22 + 32 = 0.\ \checkmark \tag{3.5.23}$$

If either had come out nonzero, one of the 2×2 determinants would contain a sign error.

Checkpoint 3.3

Two vectors lie in the xy-plane. Without any calculation, what can you say about the direction of their cross product? Could the cross product itself lie in the xy-plane?

Answer

Answers to Checkpoints

Checkpoint 3.3. The cross product is perpendicular to *both* vectors, and the only direction perpendicular to every vector in the xy-plane is along the z-axis. So the cross product points along $+z$ or $-z$ (the right-hand rule decides which). It cannot lie in the xy-plane. The one exception is trivial: if the two vectors are parallel or antiparallel, the cross product is the zero vector, which has no direction at all.

This page titled 3.5: Vector Multiplication is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.S: Vectors (Summary)

Summary

- A **scalar** has only magnitude; a **vector** has both magnitude and direction. The magnitude of $\mathbf{A}$ is written $|\mathbf{A}|$ or A, and is never negative. Two vectors are equal only if they match in both magnitude and direction; where they are drawn is irrelevant.
- A vector can be written in any of three equivalent forms: by magnitude and direction angle, by an ordered list of components $\langle A_x, A_y, A_z\rangle$, or as a sum $A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k}$.
- To **resolve** a vector into components, sketch it, build a right triangle on the acute angle you are given, compute the leg lengths $A\cos\phi$ and $A\sin\phi$, and read the signs off the sketch. Equivalently, if, and only if, θ is the standard-position angle measured counterclockwise from the $+x$-axis, then $A_x = A\cos\theta$ and $A_y = A\sin\theta$ with the signs included automatically.
- The magnitude is $A = \sqrt{A_x^2 + A_y^2}$ in two dimensions and $A = \sqrt{A_x^2 + A_y^2 + A_z^2}$ in three. The direction in two dimensions satisfies $\tan\theta = A_y/A_x$, but the calculator's $\tan^{-1}$ only returns angles from -90° to $+90^\circ$; check the signs of the components and add 180° if the result points the wrong way.
- A **unit vector** has magnitude 1 and is dimensionless. The unit vector in the direction of $\mathbf{A}$ is $\hat{A} = \mathbf{A}/A$; the basis unit vectors are $\hat{\imath}, \hat{\jmath}, \hat{k}$.
- Vectors are added and subtracted component-by-component: $(\mathbf{A} \pm \mathbf{B})_i = A_i \pm B_i$. Multiplying by a scalar c scales the magnitude by $|c|$ and reverses the direction if $c < 0$.
- The **dot product** returns a scalar: $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta = A_xB_x + A_yB_y + A_zB_z$. It is commutative and distributive, and vanishes for perpendicular vectors.
- Rearranging gives the angle between any two vectors, $\cos\theta = \dfrac{\mathbf{A}\cdot\mathbf{B}}{AB}$, and the projection of $\mathbf{A}$ onto $\mathbf{B}$, $\mathbf{A}_\parallel = (\mathbf{A}\cdot\hat{B})\hat{B}$, with the leftover piece $\mathbf{A}_\perp = \mathbf{A} - \mathbf{A}_\parallel$ perpendicular to $\mathbf{B}$.
- The **cross product** returns a vector perpendicular to both inputs: $|\mathbf{A}\times\mathbf{B}| = AB\sin\theta$, with direction given by the right-hand rule, and components most easily obtained from the determinant (3.19). It is anticommutative ($\mathbf{B}\times\mathbf{A} = -\mathbf{A}\times\mathbf{B}$) and vanishes for parallel vectors.
- $|\mathbf{A}\times\mathbf{B}|$ is the area of the parallelogram spanned by $\mathbf{A}$ and $\mathbf{B}$; half of it is the area of the triangle with those two sides.

Key Equations

Magnitude from components

$$A = \sqrt{A_x^2 + A_y^2} \quad (2\text{D}), \qquad A = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (3\text{D}) \tag{3.S.1}$$

Vector addition by components

$$\mathbf{A} + \mathbf{B} = \langle A_x + B_x,\ A_y + B_y,\ A_z + B_z\rangle \tag{3.S.2}$$

Dot product

$$\mathbf{A}\cdot\mathbf{B} = AB\cos\theta = A_xB_x + A_yB_y + A_zB_z \tag{3.S.3}$$

Angle between two vectors

$$\cos\theta = \frac{\mathbf{A}\cdot\mathbf{B}}{AB} = \frac{A_xB_x + A_yB_y + A_zB_z}{\sqrt{A_x^2 + A_y^2 + A_z^2}\,\sqrt{B_x^2 + B_y^2 + B_z^2}} \tag{3.S.4}$$

Vector projection of A onto B

$$\mathbf{A}_\parallel = (\mathbf{A}\cdot\hat{B})\,\hat{B} \tag{3.S.5}$$

Cross product magnitude

$$|\mathbf{A} \times \mathbf{B}| = AB \sin \theta \tag{3.S.6}$$

Cross product by components

$$\mathbf{A} \times \mathbf{B} = (A_y B_z - A_z B_y)\,\hat{\imath} - (A_x B_z - A_z B_x)\,\hat{\jmath} + (A_x B_y - A_y B_x)\,\hat{k} \tag{3.S.7}$$

This page titled 3.S: Vectors (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

3.E: Vectors (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 3.1 — Scalars and Vectors

3.1 ★ For each of the following quantities, state whether it is a scalar or a vector.

a. The mass of a textbook.
b. The displacement of a car from home to school.
c. The temperature outside.
d. The acceleration of a falling rock.
e. The number of pages in this book.
f. The electric force on a charged particle.
g. The speed of a car.
h. The velocity of a car.

3.2 ★ *[Conceptual.]* Answer each of the following and justify your answer in one sentence.

a. Can the magnitude of a vector ever be negative?
b. Can a component of a vector be negative?
c. Two vectors are drawn at different places on a page. What would have to be true for them to be the same vector?
d. A student says "$\mathbf{A}$ is bigger than $\mathbf{B}$." What are they probably comparing, and why is the statement sloppy as written?

3.3 ★ *[Conceptual.]* A car drives once around a circular track of radius $415\ \text{m}$ and returns to its starting point.

a. What is the magnitude of its displacement?
b. What distance did it travel?
c. Which of these two quantities is a vector, and why does the distinction matter here?

Section 3.2 — Components of a Vector

3.4 ★ A vector $\mathbf{A}$ has magnitude $12\ \text{m}$ and makes an angle of 35° above the $+x$-axis. Find its x and y components.

3.5 ★★ A force $\mathbf{F}$ has components $F_x = -6.0\ \text{N}$ and $F_y = 8.0\ \text{N}$.

a. Find the magnitude of $\mathbf{F}$.
b. Find the angle $\mathbf{F}$ makes with the $+x$-axis. *(Hint: which quadrant does $\mathbf{F}$ lie in? Use this to determine the angle.)*
c. Sketch $\mathbf{F}$ on an xy coordinate system, with the components shown.

3.6 ★★ *[Real-world: sports.]* A golf ball is struck and travels $215\ \text{m}$ at a heading 18° south of east.

a. Taking east as $+x$ and north as $+y$, what are the eastward and northward components of the ball's displacement?
b. Express the displacement in basis-vector form.
c. A second ball is hit so its displacement vector is exactly opposite: same magnitude, opposite direction. Write its displacement in component form.

3.7 ★★ *[Real-world: navigation.]* A submarine moves at $8.0\ \text{m/s}$ in a direction 40° below the horizontal, heading downward and toward the west. Take east as $+x$ and up as $+y$.

a. Sketch the velocity vector. Which quadrant does it point into, and what signs should v_x and v_y therefore have?
b. Find v_x and v_y, assigning each sign from your sketch rather than from a formula.
c. What angle does the velocity make with the $-x$-axis? Check that this is consistent with the heading you were given.

Section 3.3 — Unit Vectors

3.8 ★ Find the unit vector in the direction of each of the following.

a. $\mathbf{A} = \langle 6, 8 \rangle$
b. $\mathbf{B} = \langle -5, 12 \rangle$
c. $\mathbf{C} = 3\hat{\imath} - 4\hat{\jmath}$
d. $\mathbf{D} = \langle 1, 2, 2 \rangle$

3.9 ★★ A force $\mathbf{F}$ has magnitude $75\ \text{N}$ and points in the direction of the unit vector $\hat{n} = \langle 0.6, -0.8 \rangle$.

a. Verify that $\hat{n}$ is in fact a unit vector.
b. Find the components of $\mathbf{F}$.
c. Express $\mathbf{F}$ in basis-vector form.

3.10 ★★ A river current flows with velocity $\mathbf{v} = \langle 1.5, -2.0 \rangle$ m/s, where $+x$ is east and $+y$ is north.

a. Find the speed of the current.
b. Find the unit vector in the direction of the current.
c. In what compass direction does the current flow? Express your answer as "X degrees south of east" (or similar).

3.11 ★ *[Conceptual.]* A unit vector has a magnitude of exactly 1 and is dimensionless. When we write a velocity as $\mathbf{v} = v\ \hat{v}$, explain what role each of the two factors (v and $\hat{v}$) plays. Why is it useful to separate a vector into a magnitude and a unit vector in this way?

Section 3.4 — Vector Addition

3.12 ★ Given $\mathbf{A} = \langle 3, -2 \rangle$ and $\mathbf{B} = \langle -1, 5 \rangle$, find:

a. $\mathbf{A} + \mathbf{B}$
b. $\mathbf{A} - \mathbf{B}$
c. $|\mathbf{A} + \mathbf{B}|$
d. $2\mathbf{A} + 3\mathbf{B}$

3.13 ★★ *[Real-world: engineering.]* Three ropes pull horizontally on a stake. The first pulls with $200\ \text{N}$ to the east. The second pulls with $300\ \text{N}$ at 40° north of east. The third pulls with $250\ \text{N}$ to the south.

a. Find the net force on the stake by adding the three vectors component-by-component.
b. What is the magnitude of the net force?
c. In what direction does the net force point?
d. What single additional force would hold the stake in equilibrium? This force is called the *equilibrant*. Express your answer in component form.

3.14 ★★ *[Real-world: navigation.]* A hiker walks $2.5\ \text{km}$ due east, then turns and walks $1.8\ \text{km}$ at 35° north of east, and finally walks $1.2\ \text{km}$ due north.

a. Sketch the hiker's path on an xy diagram with east as $+x$ and north as $+y$.
b. Find the hiker's total displacement vector (in component form).
c. How far is the hiker from the starting point?
d. In what direction would the hiker have to walk in a straight line to return home?

3.15 ★★★ Three vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ sum to zero: $\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{0}$. Their magnitudes are $A = 10$, $B = 15$, and $C = 20$ (in arbitrary units). (Such a configuration is always possible whenever the magnitudes satisfy the triangle inequality.)

a. Sketch a triangle of vectors that satisfies this condition.
b. If $\mathbf{A}$ points along the $+x$-axis, find a possible set of components for $\mathbf{B}$ and $\mathbf{C}$. *(Hint: use the law of cosines, or just demand the components balance.)*

3.16 ★★ *[Conceptual.]* Two vectors $\mathbf{A}$ and $\mathbf{B}$ have fixed magnitudes A and B.

a. Under what condition is the magnitude of their sum $|\mathbf{A}+\mathbf{B}|$ as *large* as possible? What is that maximum value?
b. Under what condition is it as *small* as possible? What is that minimum value?
c. Is it possible for $|\mathbf{A}+\mathbf{B}|$ to be smaller than both A and B individually? Justify with an example or explain why not.

Section 3.5 — Vector Multiplication

3.17 ★ Compute $\mathbf{A}\cdot\mathbf{B}$ for each of the following pairs.

a. $\mathbf{A}=\langle 3,4\rangle$, $\mathbf{B}=\langle 2,-1\rangle$
b. $\mathbf{A}=2\hat{\imath}+3\hat{\jmath}$, $\mathbf{B}=-\hat{\imath}+4\hat{\jmath}$
c. $\mathbf{A}=\langle 1,-2,3\rangle$, $\mathbf{B}=\langle 4,5,-1\rangle$

3.18 ★★ Compute $\mathbf{A}\times\mathbf{B}$ for each of the following pairs.

a. $\mathbf{A}=\langle 2,0,0\rangle$, $\mathbf{B}=\langle 0,3,0\rangle$
b. $\mathbf{A}=\langle 1,2,3\rangle$, $\mathbf{B}=\langle 4,5,6\rangle$
c. $\mathbf{A}=3\hat{\imath}-2\hat{\jmath}$, $\mathbf{B}=\hat{\imath}+4\hat{\jmath}$

3.19 ★★ Two vectors have magnitudes $A=5$ and $B=8$ and make an angle of $\theta=35^\circ$ between them.

a. Find $\mathbf{A}\cdot\mathbf{B}$.
b. Find $|\mathbf{A}\times\mathbf{B}|$.
c. Verify the identity $(\mathbf{A}\cdot\mathbf{B})^2+|\mathbf{A}\times\mathbf{B}|^2=A^2B^2$ explicitly.

3.20 ★★ *[Conceptual.]* For each of the following, decide whether the two vectors are parallel, antiparallel, perpendicular, or none of these. Justify your answer using either the dot product or the cross product.

a. $\mathbf{A}=\langle 3,4\rangle$ and $\mathbf{B}=\langle 6,8\rangle$
b. $\mathbf{A}=\langle 1,2\rangle$ and $\mathbf{B}=\langle -2,-4\rangle$
c. $\mathbf{A}=\langle 1,2\rangle$ and $\mathbf{B}=\langle 4,-2\rangle$
d. $\mathbf{A}=\langle 3,1,0\rangle$ and $\mathbf{B}=\langle 1,-3,5\rangle$
e. $\mathbf{A}=\langle 2,1\rangle$ and $\mathbf{B}=\langle 3,5\rangle$
f. $\mathbf{A}=\langle 1,1,1\rangle$ and $\mathbf{B}=\langle 2,2,1\rangle$

3.21 ★★★ Given $\mathbf{A}=3\hat{\imath}-2\hat{\jmath}+\hat{k}$ and $\mathbf{B}=-\hat{\imath}+4\hat{\jmath}-2\hat{k}$:

a. Compute $\mathbf{A}\cdot\mathbf{B}$, $|\mathbf{A}|$, and $|\mathbf{B}|$.
b. Use your results and equation (3.13) to find the angle θ between $\mathbf{A}$ and $\mathbf{B}$.
c. Compute $\mathbf{A}\times\mathbf{B}$.
d. Verify your angle by computing $|\mathbf{A}\times\mathbf{B}|$ and comparing to $|\mathbf{A}||\mathbf{B}|\sin\theta$.
e. Show by direct computation that $\mathbf{A}\cdot(\mathbf{A}\times\mathbf{B})=0$. Why must this always be true?

3.22 ★★ *[Conceptual.]* The dot product of two vectors produces a scalar, while the cross product produces a vector.

a. For two nonzero vectors, when is the dot product zero? When is the magnitude of the cross product zero?
b. Can both the dot product and the cross product of two nonzero vectors be zero at the same time? Explain.
c. Without computing, explain why the expression $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is meaningful but $\mathbf{A}\cdot(\mathbf{B}\cdot\mathbf{C})$ is not.

Challenge Problems

These problems combine several ideas from the chapter and are a good test of whether the vector toolkit has become automatic.

3.23 ★★★ A force $\mathbf{F}=\langle 3,4,0\rangle$ N is applied to an object that moves along the direction $\mathbf{d}=\langle 2,1,2\rangle$. *(Use equations (3.14)–(3.16).)*

a. Find the unit vector $\hat{d}$.

b. Find the component of $\mathbf{F}$ parallel to $\mathbf{d}$, both as a scalar projection and as a vector.
c. Find the component of $\mathbf{F}$ perpendicular to $\mathbf{d}$.
d. Verify that the parallel and perpendicular pieces satisfy $F_{\parallel}^2 + F_{\perp}^2 = |\mathbf{F}|^2$.
e. Only the parallel piece does work on the object as it moves. Explain, using the geometry of Figure 3.8, why the perpendicular piece cannot.

3.24 ★★★ Three points in space are $A = (1, 0, 2)$, $B = (3, 1, 0)$, and $C = (0, 2, 1)$ (coordinates in meters).

a. Find the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$.
b. Compute $\overrightarrow{AB} \times \overrightarrow{AC}$.
c. Find the area of the triangle ABC. *(Recall equation (3.20).)*
d. Find a unit vector perpendicular to the plane of the triangle.

3.25 ★★★ *[Real-world: engineering.]* Three forces act on a bolt and hold it in equilibrium, so that $\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 = \mathbf{0}$. The first two are $\mathbf{F}_1 = \langle 4, -2, 3 \rangle$ N and $\mathbf{F}_2 = \langle -1, 5, -2 \rangle$ N.

a. Find the third force $\mathbf{F}_3$.
b. Find its magnitude.
c. What angle does $\mathbf{F}_3$ make with the z-axis? *(Hint: apply equation (3.13) with $\hat{k}$ as the second vector.)*
d. Is $\mathbf{F}_3$ perpendicular to $\mathbf{F}_1$? Justify with a dot product.

This page titled 3.E: Vectors (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4: MOTION IN TWO DIMENSIONS

In Chapter 2 we developed the language of kinematics (position, velocity, acceleration, the four equations of motion under constant acceleration) for objects moving along a single line. In Chapter 3 we built the vector tools needed to handle more than one direction. Combining these two, the description of motion in two and three dimensions is now almost effortless: each of the kinematic equations carries over from one dimension to two or three, with scalars replaced by vectors. The x, y, and z components of the motion are independent of one another, and each obeys the same equations we already know.

This independence has a powerful consequence. Consider a ball thrown forward off a tall cliff. Its horizontal motion is just constant-velocity motion in a straight line; its vertical motion is just free fall. The two happen simultaneously and independently, and the ball's actual path is the vector sum. This recognition is what allows us to solve real two-dimensional problems (projectiles, circular motion, motion in a moving frame) using the one-dimensional kinematics we already know.

Learning Objectives

- Write position, velocity, and acceleration as vector functions of time in two and three dimensions.
- Treat the components of motion along each axis independently.
- Solve projectile motion problems by analyzing horizontal and vertical motion separately.
- Use the trajectory equation to describe the shape of a projectile's path, and the range equation for launches over level ground.
- Describe uniform circular motion and compute the centripetal acceleration.
- Relate velocities and accelerations measured in different reference frames.



4.1: Kinematics in Two Dimensions

4.2: Projectile Motion

4.3: Uniform Circular Motion

4.4: Relative Motion

4.S: Motion in Two Dimensions (Summary)

4.E: Motion in Two Dimensions (Problems)

This page titled 4: Motion in Two Dimensions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.1: Kinematics in Two Dimensions

Position and displacement

In one dimension a particle's location is described by a single coordinate $x(t)$. In two dimensions we need two coordinates, $x(t)$ and $y(t)$, which together specify the tip of the **position vector**

$$\mathbf{r}(t) = x(t)\,\hat{\imath} + y(t)\,\hat{\jmath}, \tag{4.1}$$

drawn from the origin to the particle. In three dimensions we add a z-component: $\mathbf{r}(t) = x\hat{\imath} + y\hat{\jmath} + z\hat{k}$.

The **displacement** between an initial position $\mathbf{r}_1 = \langle x_1, y_1 \rangle$ and a final position $\mathbf{r}_2 = \langle x_2, y_2 \rangle$ is the vector difference

$$\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1 = (x_2 - x_1)\,\hat{\imath} + (y_2 - y_1)\,\hat{\jmath}. \tag{4.2}$$

This is just the component-by-component subtraction we learned in §3.4. In the language of Chapter 2, the displacement *along each axis* is $\Delta x = x_2 - x_1$ and $\Delta y = y_2 - y_1$; the two-dimensional displacement is the vector sum of these.

Velocity

The **average velocity** over an interval Δt is the displacement divided by the time interval:

$$\mathbf{v}_{\text{avg}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{\Delta x}{\Delta t}\,\hat{\imath} + \frac{\Delta y}{\Delta t}\,\hat{\jmath}. \tag{4.3}$$

The **instantaneous velocity** is the limit as $\Delta t \to 0$:

Key Equation: Instantaneous velocity (2D/3D)

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = v_x\hat{\imath} + v_y\hat{\jmath}, \qquad v_x = \frac{\mathrm{d}x}{\mathrm{d}t},\ v_y = \frac{\mathrm{d}y}{\mathrm{d}t}. \tag{4.4}$$

Each component of the velocity is just the time derivative of the corresponding position component. The direction of $\mathbf{v}$ at any instant is tangent to the particle's path at that instant; its magnitude $|\mathbf{v}|$ is the speed.

Acceleration

Likewise, the **average acceleration** is

$$\mathbf{a}_{\text{avg}} = \frac{\Delta\mathbf{v}}{\Delta t}, \tag{4.5}$$

and the **instantaneous acceleration** is

Key Equation: Instantaneous acceleration (2D/3D)

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = a_x\hat{\imath} + a_y\hat{\jmath}, \qquad a_x = \frac{\mathrm{d}v_x}{\mathrm{d}t},\ a_y = \frac{\mathrm{d}v_y}{\mathrm{d}t}. \tag{4.6}$$

The crucial observation is that the velocity is a vector: it can change because its magnitude changes, because its direction changes, or both. An object moving at constant speed around a circle has a nonzero acceleration, as we will see in §4.3, because its direction of motion is continuously changing.

The independence of components

For *constant acceleration*, integrating equation (4.6) twice (once to get the velocity, again to get the position) produces a set of equations identical in form to the four kinematic equations of Chapter 2, applied separately to each axis:

Key Equation: Kinematic equations for 2D constant acceleration

$$v_x = v_{0x} + a_x\,t, \qquad v_y = v_{0y} + a_y\,t, \tag{4.7}$$
$$\Delta x = v_{0x}\,t + \tfrac{1}{2}a_x t^2, \qquad \Delta y = v_{0y}\,t + \tfrac{1}{2}a_y t^2, \tag{4.8}$$
$$v_x^2 = v_{0x}^2 + 2a_x\,\Delta x, \qquad v_y^2 = v_{0y}^2 + 2a_y\,\Delta y, \tag{4.9}$$
$$\Delta x = \tfrac{1}{2}\,(v_{0x} + v_x)\,t, \qquad \Delta y = \tfrac{1}{2}\,(v_{0y} + v_y)\,t. \tag{4.10}$$

The two columns are independent: the x-motion has no idea what the y-motion is doing. The only quantity they share is the time t, which is the same in both. This single fact (that horizontal and vertical motion are independent but synchronized by a common clock) is the key to every two-dimensional kinematics problem in this chapter.

Checkpoint 4.1

A particle moves in the xy-plane with $a_x = 3.0\ \mathrm{m/s^2}$ and $a_y = 0$. At $t = 0$ it sits at the origin with velocity $\mathbf{v}_0 = \langle 0,\ 4.0\rangle\ \mathrm{m/s}$. Without doing any calculation, describe what the y-coordinate does as time passes, and what the x-coordinate does. Does either motion change what the other one is doing?

Answer

Answers to Checkpoints

Checkpoint 4.1. The y-motion has no acceleration, so y increases steadily at a constant $4.0\ \mathrm{m/s}$: $y = 4.0\ t$. The x-motion starts from rest with constant acceleration, so x grows ever faster, $x = \frac{1}{2}(3.0)t^2$. Neither motion affects the other: the x-acceleration changes only v_x, and v_y stays $4.0\ \mathrm{m/s}$ forever. The path is a parabola opening toward $+x$: a projectile "falling sideways."

This page titled 4.1: Kinematics in Two Dimensions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.2: Projectile Motion

A **projectile** is any object moving through the air with only gravity acting on it. The most common physical assumptions are that gravity is the only force (we neglect air resistance), and that the motion stays close enough to Earth's surface that $\mathbf{g}$ is a constant vector of magnitude $9.8\ \mathrm{m/s^2}$ pointing straight down. Examples: a ball thrown into the air, a cannonball fired from a cannon, or a marble that has just rolled off the edge of a table. An object becomes a projectile only once it is airborne with nothing but gravity acting on it: the marble is not a projectile while it is still on the tabletop.

With our coordinate system aligned so that $+y$ points up, the projectile's acceleration is

$$\mathbf{a} = \langle 0,\ -g \rangle, \tag{4.2.1}$$

i.e. $a_x = 0$ and $a_y = -g$. The horizontal motion is uniform (constant velocity); the vertical motion is free fall with constant downward acceleration g.

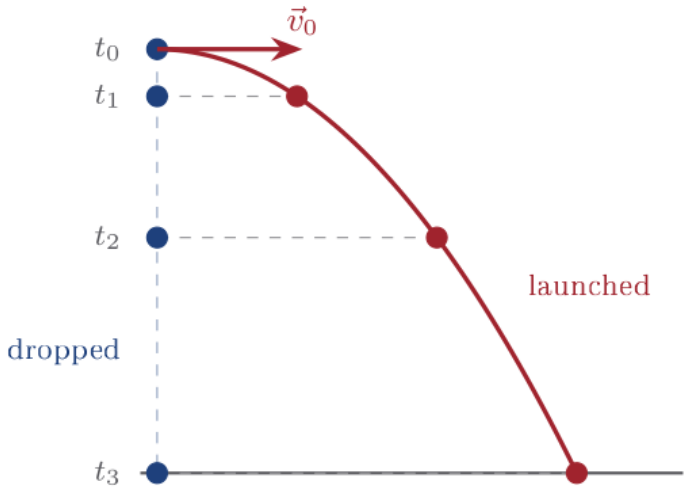


Figure 4.1. Two balls released at the same instant, photographed at equal time intervals. The blue ball is dropped from rest; the red ball is launched horizontally. The horizontal motion of the red ball has no effect whatsoever on its vertical motion, so at every instant the two balls are at the same height (dashed lines) and they reach the ground together. Notice also that the red ball's horizontal spacing is uniform (constant horizontal velocity) while its vertical spacing grows, because the vertical motion is accelerating.

•Figure description

Two balls released simultaneously and shown at equal time intervals. The blue ball is dropped straight down; the red ball is launched horizontally and follows a curved path. At every instant both balls are at the same height, showing that horizontal motion does not affect vertical fall.

Figure 4.1 shows what the independence of the two motions looks like in practice. Two balls leave the same point at the same instant: one is simply dropped, the other is launched horizontally. Because neither ball has any vertical velocity to begin with, and because the horizontal motion of the second ball does nothing at all to its vertical motion, the two fall at exactly the same rate. At every instant they are at the same height, and they land together.

Solving projectile problems

Every projectile problem comes down to writing the two columns of equations, (4.7) through (4.10), with $a_x = 0$ and $a_y = -g$, and identifying which quantities are known and which are sought. A standard strategy:

1. Choose a coordinate system. Usually $+x$ is the direction of horizontal motion and $+y$ is up. This choice is completely up to the person doing the problem. There is no wrong choice so long as the axes are perpendicular to each other and stay fixed for the whole problem. You can just as easily let the $+y$ direction be down. As long as you are consistent, your answer will be the same.

2. Decompose the initial velocity into components.
3. Write down the kinematic equations for each axis independently.
4. Identify the link between the x and y motion: almost always, the elapsed time t.
5. Solve.

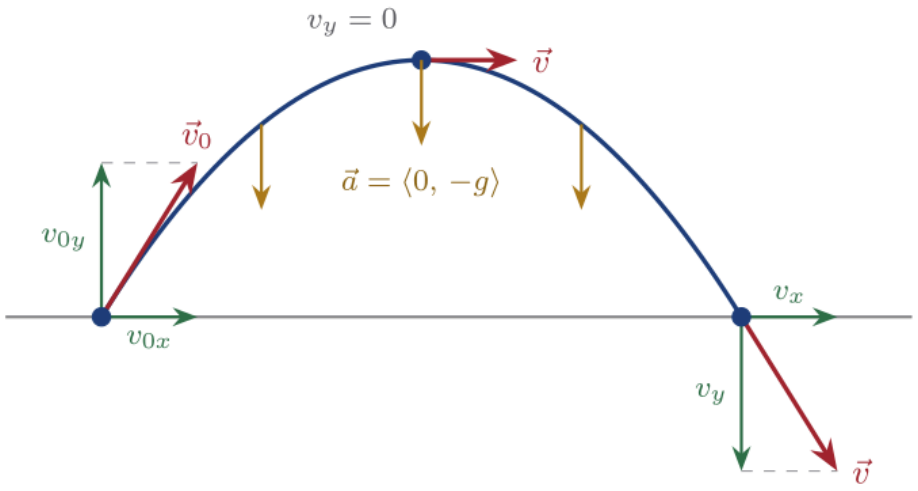


Figure 4.2. Velocity and acceleration during a projectile flight. The horizontal component v_x (green) is the same at every point. The vertical component v_y starts positive, shrinks to zero at the top, then grows negative, and on level ground the landing speed equals the launch speed. The acceleration (gold) is the same constant downward vector at every instant of the flight, including at the very top.

•Figure description

A projectile's parabolic trajectory with velocity vectors drawn at several points. The green horizontal component stays the same length throughout. The vertical component starts upward, shrinks to zero at the peak where v sub y equals zero, then grows downward. Acceleration points downward everywhere.

Note: The sign of a_y — and what happens if you point $+y$ down

Every equation in this chapter is written for the choice $+y$ up, which makes $a_y = -g$, where $g = 9.8\ \text{m/s}^2$ is a *positive* number. If you instead choose $+y$ pointing down (often the natural choice when something simply falls from a height), then $a_y = +g$, and downward displacements and velocities all come out positive.

Neither choice is more correct, and both give the same physical answer. What is not allowed is mixing them: declaring $+y$ up and then writing $\Delta y = +14\ \text{m}$ for a rock that fell $14\ \text{m}$. Nearly every sign error in projectile motion traces back to that one slip. Draw your axes, write the direction next to them, and hold to it for the whole problem.

Figure 4.2 summarizes what the velocity and acceleration vectors are doing over the course of a flight.

Note: At the top of the arc, the acceleration is *not* zero

At the highest point of the flight the vertical velocity passes through zero: the projectile has stopped rising and has not yet begun to fall. It is tempting to conclude that the acceleration must be zero there too. It is not. The acceleration is g, straight down, at every instant of the flight, apex included.

The test is easy to apply. If the acceleration really were zero at the top, then v_y would have no reason to change, and the projectile would sail off horizontally and never come down. What is momentarily zero at the apex is v_y, not a_y, and a quantity can be zero at an instant while changing as fast as ever.

Example 4.1: A baseball off the bat

A baseball is struck and departs from the bat horizontally with a speed of $41.2\ \text{m/s}$. The ball strikes the ground a horizontal distance of $19.5\ \text{m}$ from the bat. How far above the field was the baseball when it left the bat?

Solution

Take $+x$ in the direction of motion and $+y$ up, with the origin at the point of release. The initial velocity is purely horizontal:

$$v_{0x} = 41.2\,\text{m/s}, \qquad v_{0y} = 0. \tag{4.2.2}$$

The acceleration is $a_x = 0$, $a_y = -g$.

Step 1: Decide which direction to start with. The vertical equation contains two unknowns, the drop Δy and the time t, so it cannot be solved yet. The horizontal equation contains only one unknown, t, because we know both Δx and v_{0x}. Start there.

Step 2: Time of flight from the horizontal motion. Since $a_x = 0$, the horizontal motion is at constant velocity, $\Delta x = v_{0x}\ t$. Divide both sides by v_{0x} to isolate t:

$$t = \frac{\Delta x}{v_{0x}} = \frac{19.5\,\text{m}}{41.2\,\text{m/s}} \approx 0.473\,\text{s}. \tag{4.2.3}$$

Step 3: Drop from the vertical motion. The ball is in the air for one and the same time interval whichever direction we look at, so this t goes straight into the vertical equation. With $v_{0y} = 0$ the first term vanishes:

$$\Delta y = v_{0y}\,t + \tfrac{1}{2}a_y t^2 = 0 + \tfrac{1}{2}(-9.8\,\text{m/s}^2)(0.473\,\text{s})^2 \approx -1.10\,\text{m}. \tag{4.2.4}$$

Only the time is squared, not the -9.8: $(0.473)^2 = 0.224\ \text{s}^2$, and $\frac{1}{2}(-9.8)(0.224) = -1.10$. The units work out as $(\text{m/s}^2)(\text{s}^2) = \text{m}$.

The negative sign indicates the ball drops $1.10\ \text{m}$ during its flight. Since it started at $y = 0$ and ended on the ground, the ball was approximately $1.10\ \text{m}$ above the field when it left the bat.

Example 4.2: Throwing from a hot-air balloon

A rock is thrown from a hot-air balloon that is hovering at rest $14\ \text{m}$ above the ground. The rock leaves the thrower's hand at an angle of 24° below the horizontal, with an initial speed of $18\ \text{m/s}$.

a. How long is the rock in the air?
b. How far from the launch point (horizontally) does the rock land?

Solution

Place the origin at the launch point with $+x$ in the direction of the throw and $+y$ up. Then the ground is at $y = -14\ \text{m}$.

Components of the initial velocity. An angle *below* the horizontal, measured from $+x$, is a clockwise angle, so its standard-position value is -24°. Using it in equations (3.1) and (3.2) lets the trigonometric functions supply the signs:

$$v_{0x} = (18)\cos(-24^\circ) \approx +16.4\,\text{m/s}, \qquad v_{0y} = (18)\sin(-24^\circ) \approx -7.32\,\text{m/s}. \tag{4.2.5}$$

The cosine of a negative angle is positive and the sine is negative, which matches the sketch: the rock starts out moving forward and downward. (Building the right triangle on the 24° angle and attaching the signs by eye, as in Chapter 3, gives the same two numbers.)

(a) The rock lands when $\Delta y = -14\ \text{m}$. Using $\Delta y = v_{0y}t + \frac{1}{2}(-g)t^2$, and dropping units for clarity (t in seconds, lengths in meters):

$$-14 = -7.32\,t - \tfrac{1}{2}(9.8)t^2 = -7.32\,t - 4.9\,t^2. \tag{4.2.6}$$

This contains both t and t^2, so it is a quadratic. Move every term to the left-hand side by adding $4.9\ t^2$ and $7.32\ t$ to both sides:

$$4.9\,t^2 + 7.32\,t - 14 = 0. \tag{4.2.7}$$

Read off the coefficients with their signs: $a = 4.9$, $b = 7.32$, $c = -14$. In the quadratic formula,

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{4.2.8}$$

the term $-4ac = -4(4.9)(-14) = +274.4$ is positive because c is negative. So

$$t = \frac{-7.32 \pm \sqrt{(7.32)^2 + 274.4}}{2(4.9)} = \frac{-7.32 \pm \sqrt{53.6 + 274.4}}{9.8} = \frac{-7.32 \pm 18.11}{9.8}. \tag{4.2.9}$$

The "plus" root gives $t = 10.79/9.8 \approx 1.10$ s. The "minus" root gives $t = -25.43/9.8 \approx -2.59$ s, a time before the rock was thrown, so it is discarded.

(b) Horizontally there is no acceleration, so the rock covers ground at the constant rate v_{0x} for the whole 1.10 s:

$$\Delta x = v_{0x}\, t = (16.4\,\text{m/s})(1.10\,\text{s}) \approx 18.1\,\text{m}. \tag{4.2.10}$$

The rock lands about 18 m from a point directly below the balloon.

Check. Substitute $t = 1.10$ s back into the vertical equation: $-7.32(1.10) - 4.9(1.10)^2 = -8.05 - 5.93 = -13.98$ m ≈ -14 m, the height of the balloon. ✓

Example 4.3: Bowling ball off a bridge

A bowling ball rolls horizontally off a bridge 6.0 m above the water below. It lands 2.5 m horizontally from the point directly below where it left the bridge.

a. With what speed did the ball leave the bridge?
b. With what velocity (magnitude and direction) does it hit the water?

Solution

Origin at the point of departure, $+x$ horizontal in the direction of motion, $+y$ up. The initial velocity is purely horizontal: $v_{0x} = v_0$, $v_{0y} = 0$. The water is at $y = -6.0$ m.

(a) This time the horizontal equation has two unknowns (v_0 and t), but the vertical one has only t, so start with the vertical motion. With $v_{0y} = 0$, $\Delta y = -\frac{1}{2}gt^2$:

$$-6.0 = -\tfrac{1}{2}(9.8)t^2 = -4.9\,t^2. \tag{4.2.11}$$

Divide both sides by -4.9 (the minus signs cancel), then take the positive square root, since elapsed time is positive:

$$t^2 = \frac{6.0}{4.9} = 1.224\,\text{s}^2 \quad \Longrightarrow \quad t = \sqrt{1.224\,\text{s}^2} \approx 1.107\,\text{s}. \tag{4.2.12}$$

Now the horizontal motion has only one unknown. Solve $\Delta x = v_0 t$ for v_0 by dividing both sides by t:

$$v_0 = \frac{\Delta x}{t} = \frac{2.5\,\text{m}}{1.107\,\text{s}} \approx 2.26\,\text{m/s}. \tag{4.2.13}$$

(b) A velocity has two components, and each is found separately. Horizontally nothing changes, so just before landing $v_x = v_0 = 2.26$ m/s. Vertically, equation (2.8) gives

$$v_y = v_{0y} + a_y t = 0 + (-9.8\,\text{m/s}^2)(1.107\,\text{s}) \approx -10.85\,\text{m/s}. \tag{4.2.14}$$

The two components are the legs of a right triangle whose hypotenuse is the speed:

$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{(2.26)^2 + (-10.85)^2}\,\text{m/s} = \sqrt{5.1 + 117.7}\,\text{m/s} \approx 11.1\,\text{m/s}. \tag{4.2.15}$$

For the direction, the angle below the horizontal is the angle whose opposite side is $|v_y|$ and adjacent side is v_x:

$$\theta = \tan^{-1}\left(\frac{|v_y|}{v_x}\right) = \tan^{-1}\left(\frac{10.85}{2.26}\right) \approx \tan^{-1}(4.80) \approx 78^\circ \tag{4.2.16}$$

below the horizontal. Using $|v_y|$ gives the size of the angle; the sketch supplies "below." The ball hits the water moving fast and steeply downward.

Check. The vertical speed can also be found without the time, from $v_y^2 = v_{0y}^2 + 2a_y\Delta y = 0 + 2(-9.8)(-6.0) = 117.6\ \text{m}^2/\text{s}^2$, giving $|v_y| = 10.8\ \text{m/s}$: the same answer, from the given data alone.

Example 4.4: Jumping between buildings

You are stranded on top of a burning building. The nearest safe building is $6.50\ \text{m}$ away horizontally and $1.00\ \text{m}$ lower than where you stand. You decide to run and jump, leaving the edge of your roof at $8.10\ \text{m/s}$ at an angle of 35° above the horizontal. Do you make it across?

Solution

Origin at the take-off point, $+x$ horizontal toward the safe building, $+y$ up. Initial velocity components:

$$v_{0x} = (8.10)\cos 35^\circ \approx 6.63\,\text{m/s}, \qquad v_{0y} = (8.10)\sin 35^\circ \approx 4.65\,\text{m/s}. \tag{4.2.17}$$

The edge of the safe building is at $x = 6.50\ \text{m}$, and its roof is at $y = -1.00\ \text{m}$.

Strategy. "Do you make it?" becomes a precise question: at the instant you are directly above the far edge, are you above or below its roof? So find *when* you reach $x = 6.50\ \text{m}$, then find *where* you are vertically at that instant.

Step 1: When do you reach the far edge? Horizontally the velocity is constant, so $\Delta x = v_{0x}\ t$. Divide both sides by v_{0x}:

$$t = \frac{\Delta x}{v_{0x}} = \frac{6.50\,\text{m}}{6.63\,\text{m/s}} \approx 0.980\,\text{s}. \tag{4.2.18}$$

Step 2: Where are you vertically at that instant? Substitute this time into the vertical equation, with $a_y = -g$:

$$\Delta y = v_{0y}\,t - \tfrac{1}{2}g\,t^2 = (4.65)(0.980) - \tfrac{1}{2}(9.8)(0.980)^2 \approx 4.56 - 4.71 \approx -0.15\,\text{m}. \tag{4.2.19}$$

The first term is how high your upward launch would carry you with no gravity; the second is how far gravity pulls you back down in the same time. Gravity wins, slightly, so you are $0.15\ \text{m}$ below your starting height.

Step 3: Compare. The far roof is at $-1.00\ \text{m}$, and $-0.15\ \text{m}$ is higher than $-1.00\ \text{m}$ (it is less negative). You clear the edge with $-0.15 - (-1.00) = 0.85\ \text{m}$ to spare, so yes, you make it.

The shape of the path: trajectory and range

The two columns of kinematic equations give x and y separately as functions of time. Sometimes we want the *shape* of the path itself: y as a function of x, with the time eliminated.

Take a launch from the origin with speed v_0 at an angle θ above the horizontal, so that $v_{0x} = v_0\cos\theta$ and $v_{0y} = v_0\sin\theta$. The horizontal equation gives the time in terms of how far the projectile has traveled horizontally,

$$x = (v_0\cos\theta)\,t \quad \Longrightarrow \quad t = \frac{x}{v_0\cos\theta}, \tag{4.2.20}$$

and substituting that into the vertical equation $y = (v_0\sin\theta)\ t - \frac{1}{2}gt^2$ gives

Key Equation: Trajectory of a projectile

$$y = x\tan\theta - \frac{g\,x^2}{2v_0^2}\sec^2\theta, \tag{4.11}$$

where $\sec^2\theta = 1/\cos^2\theta$. Since v_0, θ, and g are all constants, this has the form $y = Ax - Bx^2$ with $B > 0$: the path of a projectile is a downward-opening **parabola**, which is exactly the curve drawn in Figures 4.1 and 4.2.

If the projectile lands at the same height it started from (level ground), we can find how far it goes by setting $y = 0$ and solving equation (4.11) for x. Factoring out the $x = 0$ root (the launch point itself) leaves the **range**:

Key Equation: Range on level ground

$$R = \frac{v_0^2\sin 2\theta}{g}. \tag{4.12}$$

Two consequences are worth noticing. First, $\sin 2\theta$ is largest when $2\theta = 90^\circ$, so on level ground the greatest range comes from a launch at $\theta = 45^\circ$. Second, $\sin 2\theta$ takes the same value at θ and at $90^\circ - \theta$, so *complementary launch angles give the same range*: 30° and 60° carry a projectile equally far, as do 40° and 50°. The low shot gets there faster and flatter; the high shot spends longer in the air but covers less ground per second. They arrive at the same place.

Note: When equation (4.12) does *not* apply

Equation (4.12) is a shortcut, and like most shortcuts it comes with fine print: it was derived by assuming the projectile lands at exactly the same height it was launched from. It does not apply to a ball hit off a bat and caught on the ground, to a rock thrown from a balloon, or to any of the cliff problems in this chapter, and using it there is one of the most common ways to get a projectile problem wrong.

Equation (4.11), on the other hand, holds for any launch from the origin. When the landing height differs from the launch height, go back to the two columns of kinematic equations and solve for the time of flight; that method always works.

Checkpoint 4.2

Two objects are launched from the edge of a tall cliff at the same instant. One is dropped from rest; the other is thrown horizontally with speed $15\ \mathrm{m/s}$. Which hits the ground first? Justify your answer using the independence of horizontal and vertical motion.

Answer

Answers to Checkpoints

Checkpoint 4.2. They hit the ground at the same instant. Both start with zero vertical velocity and both have the same vertical acceleration, g downward, so their vertical motions are identical at every moment. The thrown object's $15\ \mathrm{m/s}$ is entirely horizontal and has no effect on how fast it falls; it only makes the object land farther from the base of the cliff.

This page titled 4.2: Projectile Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.3: Uniform Circular Motion

Uniform circular motion is motion along a circular path at constant speed. In physics, “uniform” means “constant in magnitude”, so the speed is constant, but the velocity is not, because the direction of motion is continuously changing. A particle moving in a circle is therefore always accelerating, even at constant speed.

Centripetal acceleration

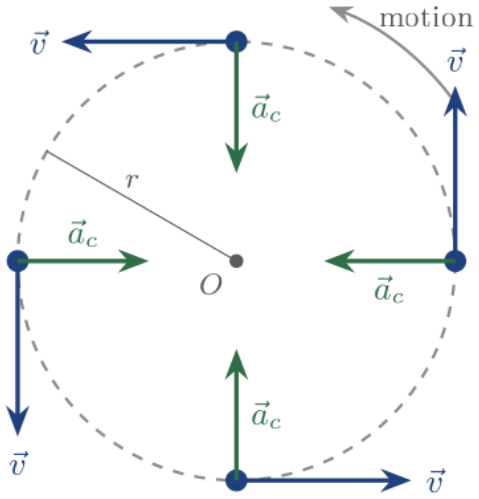


Figure 4.3. In uniform circular motion the velocity (blue) is always tangent to the circle, and the centripetal acceleration (green) always points toward the center. Both vectors keep a constant magnitude while their directions change continuously. The two arrow lengths are not to a common scale: velocity and acceleration are different quantities and cannot be compared on the page.

•Figure description

A particle moving counterclockwise on a circle of radius r centered at O. A blue velocity arrow is tangent to the circle and a green acceleration arrow points inward toward the center. Both keep constant magnitude while continuously changing direction.

If the speed is constant, the only thing the velocity can change is its direction, and the direction is always tangent to the circle. Over a short time interval the velocity direction turns slightly, so the change $\Delta\mathbf{v}$ points into the circle. In the limit $\Delta t \to 0$ that change becomes exactly perpendicular to the velocity, and the acceleration points straight at the center; see Figure 4.3. (Over a *finite* interval $\Delta\mathbf{v}$ only points roughly inward: it is the limit that makes it exact.) The acceleration of a particle in uniform circular motion is called the **centripetal acceleration**, from the Latin for “center seeking.”

We can get the size of that acceleration directly. Put the center of the circle at the origin, let the particle start at $(r, 0)$, and let it travel counterclockwise. Write T for the time it takes to go once around, the **period**, which we look at more closely in §4.3. Because the speed is constant, the particle sweeps out equal angles in equal times, so after a time t it has swept through the fraction t/T of a full turn, an angle of $2\pi t/T$. Its position is therefore

$$\mathbf{r}(t) = r\cos\left(\frac{2\pi t}{T}\right)\hat{\imath} + r\sin\left(\frac{2\pi t}{T}\right)\hat{\jmath}. \tag{4.3.1}$$

Differentiating once with the chain rule gives the velocity, and differentiating again gives the acceleration:

$$\begin{aligned} \mathbf{v}(t) &= \frac{2\pi r}{T}\left[-\sin\left(\frac{2\pi t}{T}\right)\hat{\imath} + \cos\left(\frac{2\pi t}{T}\right)\hat{\jmath}\right], \\ [4pt]\mathbf{a}(t) &= -\left(\frac{2\pi}{T}\right)^2\left[r\cos\left(\frac{2\pi t}{T}\right)\hat{\imath} + r\sin\left(\frac{2\pi t}{T}\right)\hat{\jmath}\right] = -\left(\frac{2\pi}{T}\right)^2\mathbf{r}(t). \end{aligned} \tag{4.3.2}$$

Three things fall out at once. The speed is $|\mathbf{v}| = 2\pi r/T$, the circumference divided by the period, which is what it had better be. The acceleration is a *negative* multiple of the position vector, so it points from the particle straight back at the center, confirming what Figure 4.3 shows. And its magnitude is

$$a_c = \left(\frac{2\pi}{T}\right)^2 r = \frac{(2\pi r/T)^2}{r} = \frac{v^2}{r}, \tag{4.3.3}$$

which is the result we were after:

Key Equation: Centripetal acceleration

$$a_c = \frac{v^2}{r}, \tag{4.13}$$

where v is the (constant) speed and r is the radius of the circular path. The acceleration always points from the particle's current position toward the center of the circle.

Note that the speed enters *squared*. Doubling the speed around the same curve does not double the acceleration: it quadruples it. This is why a curve that feels comfortable at $30\ \mathrm{km/hr}$ can be genuinely dangerous at $60\ \mathrm{km/hr}$, and it is worth committing to memory before Chapter 5, where we ask what force is required to produce this acceleration.

Note: "Centripetal" names a direction, not a force

Centripetal means "directed toward the center." It tells you where the acceleration points; it says nothing about what causes it. When we take up forces in Chapter 5, the answer to "what pushes the object toward the center?" will be different every time: tension in a string, friction between tire and road, gravity on a satellite, the normal force from a banked track.

There is no separate "centripetal force" in nature waiting to be added to your list of forces. The phrase is shorthand for whichever ordinary force happens to be pointing at the center in the problem at hand.

Checkpoint 4.3

Two cars round two different circular curves, each at constant speed. Car A takes a curve of radius r at speed v. Car B takes a curve of radius $2r$ at speed $2v$. Which car has the larger centripetal acceleration, and by what factor?

Answer

Period and frequency

A useful auxiliary quantity is the **period of revolution**, T: the time it takes the particle to complete one full circle. Since one circle has circumference $2\pi r$ and the particle moves at constant speed v,

Key Equation: Period of uniform circular motion

$$T = \frac{2\pi r}{v}. \tag{4.14}$$

The **frequency** f counts revolutions per unit time, and is measured in hertz ($1\ \mathrm{Hz} = 1\ \mathrm{s}^{-1}$) or, often in mechanics, in revolutions per minute (rpm).

Key Equation: Frequency

$$f = \frac{1}{T} \tag{4.15}$$

Example 4.5: Artificial gravity on a spacecraft

A long-duration mission to Mars would require some form of artificial gravity to keep the astronauts healthy. One method is to spin a section of the spacecraft so that the centripetal acceleration felt by the crew matches the gravitational acceleration on Earth. For a circular section of radius $12\ \text{m}$, how many revolutions per minute are required to simulate Earth gravity?

Solution

Step 1: Find the required speed. We want $a_c = g = 9.8\ \text{m/s}^2$. Set equation (4.13) equal to g and solve for v: multiply both sides by r, then take the square root of both sides:

$$\frac{v^2}{r} = g \quad \Longrightarrow \quad v^2 = gr \quad \Longrightarrow \quad v = \sqrt{gr}, \tag{4.3.4}$$

so

$$v = \sqrt{(9.8\,\text{m/s}^2)(12\,\text{m})} = \sqrt{117.6\,\text{m}^2/\text{s}^2} \approx 10.84\,\text{m/s}. \tag{4.3.5}$$

Step 2: Find the period. One revolution covers the circumference $2\pi r$ at speed v, so by equation (4.14)

$$T = \frac{2\pi r}{v} = \frac{2\pi(12\,\text{m})}{10.84\,\text{m/s}} = \frac{75.4\,\text{m}}{10.84\,\text{m/s}} \approx 6.95\,\text{s}. \tag{4.3.6}$$

Step 3: Convert to revolutions per minute. By equation (4.15), $f = 1/T$ is revolutions per *second*:

$$f = \frac{1\,\text{rev}}{6.95\,\text{s}} = 0.144\,\text{rev/s}. \tag{4.3.7}$$

There are 60 seconds in a minute, so multiply by $60\ \text{s}/1\ \text{min}$ to cancel the seconds:

$$f = 0.144\,\frac{\text{rev}}{\cancel{\text{s}}} \times \frac{60\,\cancel{\text{s}}}{1\,\text{min}} \approx 8.6\,\text{rpm}. \tag{4.3.8}$$

A shortcut worth seeing. Substituting $v = \sqrt{gr}$ into $T = 2\pi r/v$ gives $T = 2\pi r/\sqrt{gr} = 2\pi\sqrt{r/g}$, since $r/\sqrt{r} = \sqrt{r}$. Then $T = 2\pi\sqrt{12/9.8} = 6.95\ \text{s}$ in one step. The formula also shows at a glance that a larger radius means a longer period: a slower spin.

A rotation rate of about $8.6\ \text{rpm}$ is needed. That is uncomfortably fast: early studies of rotating habitats recommended staying near $2\ \text{rpm}$ so that crew members would not get dizzy every time they turned their heads, although more recent work suggests people can adapt to somewhat higher rates. Either way, a larger radius gives the same a_c at a slower spin, which is why proposed rotating habitats tend to be far bigger than $12\ \text{m}$ in radius.

Example 4.6: A pilot's banked turn

A pilot enters a horizontal circular turn with initial velocity $\mathbf{v}_i = (400\ \hat{\imath} + 500\ \hat{\jmath})\ \text{m/s}$, and after $24.0\ \text{s}$ leaves the turn with velocity $\mathbf{v}_f = (-400\ \hat{\imath} - 500\ \hat{\jmath})\ \text{m/s}$. What is the magnitude of the centripetal acceleration during the turn, in units of g?

Solution

Step 1: Read the geometry from the velocities. The speed is the magnitude of the velocity, found from its components:

$$|\mathbf{v}_i| = \sqrt{400^2 + 500^2}\,\text{m/s} = \sqrt{160\,000 + 250\,000}\,\text{m/s} = \sqrt{410\,000}\,\text{m/s} \approx 640\,\text{m/s}, \tag{4.3.9}$$

and $|\mathbf{v}_f|$ is the same, since squaring removes the minus signs. So the speed is constant, as it must be in uniform circular motion. Moreover $\mathbf{v}_f = -\mathbf{v}_i$: the plane ends up flying in exactly the opposite direction. The velocity is always tangent to the circle, so reversing it means the plane has gone halfway around: a 180° turn.

Step 2: Find the radius. If half a revolution takes $24.0\ \text{s}$, then the full period is $T = 2(24.0\ \text{s}) = 48.0\ \text{s}$. Solve equation (4.14), $T = 2\pi r/v$, for r by multiplying both sides by v and dividing both sides by 2π:

$$r = \frac{vT}{2\pi} = \frac{(640\,\text{m/s})(48.0\,\text{s})}{2\pi} \approx 4890\,\text{m}. \tag{4.3.10}$$

Step 3: Find the acceleration.

$$a_c = \frac{v^2}{r} = \frac{(640\,\text{m/s})^2}{4890\,\text{m}} \approx 83.8\,\text{m/s}^2. \tag{4.3.11}$$

The radius was only a stepping stone. Substituting $r = vT/(2\pi)$ into $a_c = v^2/r$ cancels one factor of v, leaving $a_c = 2\pi v/T = 2\pi(640)/48.0 = 83.8\ \text{m/s}^2$ with no intermediate rounding: a useful check.

Step 4: Express it in g's. Divide by g; the units cancel and leave a pure number:

$$\frac{a_c}{g} = \frac{83.8\ \cancel{\text{m/s}^2}}{9.8\ \cancel{\text{m/s}^2}} = 8.55 \approx 8.6. \tag{4.3.12}$$

The pilot experiences about $8.6\ g$. This is an extreme maneuver. Trained pilots in g-suits tolerate roughly $9\ g$, but only for a few seconds at a time; holding $8.6\ g$ for a full $24\ \text{s}$ is well beyond human endurance, and an unprotected passenger would lose consciousness almost at once. The numbers here describe the geometry of the turn, not a flight anyone would walk away from.

Answers to Checkpoints

Checkpoint 4.3. Car B. Its centripetal acceleration is $a_B = (2v)^2/(2r) = 4v^2/(2r) = 2\ v^2/r$, while car A's is $a_A = v^2/r$. So car B's acceleration is larger by a factor of 2: doubling the speed quadruples v^2, and doubling the radius only halves the result.

This page titled 4.3: Uniform Circular Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.4: Relative Motion

All motion is relative. When we say a car is moving at $30\ \mathrm{m/s}$, we always mean *relative to something else*: in this case, the ground. The choice of what to measure motion against is called the choice of **reference frame**. Different reference frames assign different velocities and (in general) different accelerations to the same particle.

The velocity-addition formula

Suppose an object P (a particle, a vehicle, an animal, anything moving) is observed from two different reference frames: frame A and frame B. Frame B is itself moving relative to frame A. Let

- $\mathbf{v}_{PA}$ = velocity of P as measured in frame A
- $\mathbf{v}_{PB}$ = velocity of P as measured in frame B
- $\mathbf{v}_{BA}$ = velocity of frame B as measured in frame A

Then the three are related by simple vector addition:

Key Equation: Relative velocity (Galilean transformation)

$$\mathbf{v}_{PA} = \mathbf{v}_{PB} + \mathbf{v}_{BA}. \tag{4.16}$$

A useful mnemonic: the inner subscripts ("B") on the right-hand side cancel, leaving the outer subscripts ("PA"), exactly what appears on the left-hand side.

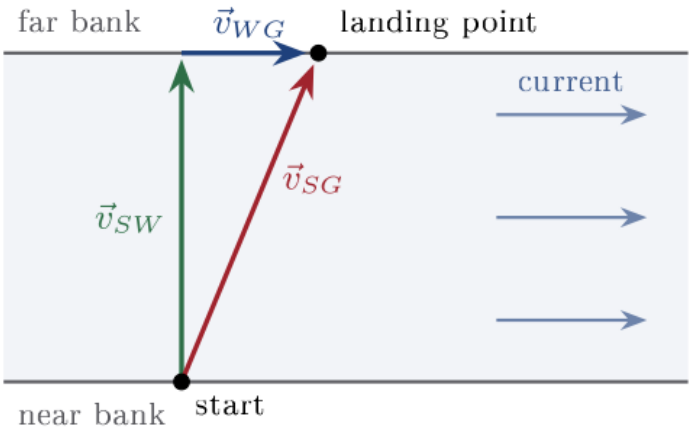


Figure 4.4. A swimmer (S) who points herself straight across a river swims at $\mathbf{v}_{SW}$ relative to the water (W), while the water moves at $\mathbf{v}_{WG}$ relative to the ground (G). Her velocity relative to the ground is the vector sum $\mathbf{v}_{SG} = \mathbf{v}_{SW} + \mathbf{v}_{WG}$: inner subscripts W cancel, outer subscripts SG survive. Because the current has no component *across* the river, she reaches the far bank in exactly the time it would take her in still water; the current only decides where along the bank she arrives.

•Figure description

A river flowing between a near bank and a far bank. A swimmer aims straight across with velocity relative to the water, while the current carries her downstream. The two vectors add to give a resultant path angled downstream, so she lands below her starting point on the opposite bank.

Figure 4.4 shows the formula at work in the situation it is most often needed for: crossing a moving river.

This formula generalizes the intuitive statement "your speed relative to the ground equals your speed relative to the train plus the train's speed relative to the ground." It applies equally to a swimmer in a current, an airplane in a headwind, or a passenger walking inside a moving subway car.

Example 4.7: Cars on the highway

You are driving at $80\ \text{mi/hr}$ on a straight freeway. Another car, in the same lane in front of you, is moving at $70\ \text{mi/hr}$ in the same direction. How fast does the other car appear to move, in your reference frame?

Solution

Let P be the other car, A be the ground frame, and B be your car's frame. Then $\mathbf{v}_{PA} = 70\ \hat{\imath}$ mi/hr, and $\mathbf{v}_{BA} = 80\ \hat{\imath}$ mi/hr. We want $\mathbf{v}_{PB}$, the velocity of the other car relative to you. Equation (4.16) says $\mathbf{v}_{PA} = \mathbf{v}_{PB} + \mathbf{v}_{BA}$; subtract $\mathbf{v}_{BA}$ from both sides to isolate $\mathbf{v}_{PB}$:

$$\mathbf{v}_{PB} = \mathbf{v}_{PA} - \mathbf{v}_{BA} = (70 - 80)\,\hat{\imath}\,\text{mi/hr} = -10\,\hat{\imath}\,\text{mi/hr}. \tag{4.4.1}$$

In your frame the other car drifts backward toward you (the $-\hat{\imath}$ direction) at $10\ \text{mi/hr}$, which is exactly what you see through the windshield as you slowly close the gap and overtake it.

Acceleration in different reference frames

If frame B moves with constant velocity relative to frame A (i.e., $\mathbf{v}_{BA}$ is constant), then differentiating equation (4.16) gives

$$\mathbf{a}_{PA} = \mathbf{a}_{PB}. \tag{4.17}$$

The acceleration of P is the same in both frames. Two observers in uniform relative motion will always agree on a particle's acceleration, even though they disagree on its velocity. This fact will be crucial when we introduce Newton's laws in Chapter 5: any frame moving at constant velocity relative to a frame in which Newton's laws hold is itself a frame in which Newton's laws hold.

If B is itself accelerating relative to A, equation (4.17) acquires an extra term $\mathbf{a}_{BA}$, and the two observers disagree on the acceleration as well.

Example 4.8: A camera operator and a cheetah

A camera operator on a pickup truck is traveling westward at $20\ \text{km/hr}$ while recording a cheetah running westward $30\ \text{km/hr}$ faster than the truck. The cheetah stops, turns, and then runs at $45\ \text{km/hr}$ eastward, as measured by a nervous crew member who stands beside the cheetah's path. The change in the animal's velocity takes $2.0\ \text{s}$.

a. What is the magnitude and direction of the cheetah's acceleration as measured by the ground-based crew member?
b. What is the magnitude and direction of the cheetah's acceleration as measured by the camera operator on the moving truck?

Solution

Take $+x$ as east, so every westward velocity is negative. The truck moves at $-20\ \hat{\imath}\ \text{km/hr}$ relative to the ground.

Initial cheetah velocity, ground frame: westward and $30\ \text{km/hr}$ faster than the truck, so its speed is $20 + 30 = 50\ \text{km/hr}$ and its velocity is $\mathbf{v}_{C,\text{initial}} = -50\ \hat{\imath}\ \text{km/hr}$.

Final cheetah velocity, ground frame: $+45\ \hat{\imath}\ \text{km/hr}$ (eastward).

(a) Ground-frame acceleration. The change in velocity is final minus initial, and subtracting a negative velocity adds its size:

$$\mathbf{a}_{\text{ground}} = \frac{\Delta \mathbf{v}}{\Delta t} = \frac{45 - (-50)}{2.0}\,\hat{\imath}\,\text{km/(hr}\cdot\text{s)} = \frac{95}{2.0}\,\hat{\imath}\,\text{km/(hr}\cdot\text{s)} = 47.5\,\hat{\imath}\,\text{km/(hr}\cdot\text{s)}. \tag{4.4.2}$$

The cheetah's velocity changes by $95\ \text{km/hr}$, not $5\ \text{km/hr}$, because it first has to lose $50\ \text{km/hr}$ of westward speed and then gain $45\ \text{km/hr}$ of eastward speed.

The unit "km/hr per second" is awkward, so convert. Since $1\ \text{km} = 1000\ \text{m}$ and $1\ \text{hr} = 3600\ \text{s}$,

$$1\,\frac{\text{km}}{\text{hr}} = \frac{1000\,\text{m}}{3600\,\text{s}} = \frac{1}{3.6}\,\frac{\text{m}}{\text{s}}, \tag{4.4.3}$$

so dividing a speed in km/hr by 3.6 converts it to m/s. Then

$$\mathbf{a}_{\text{ground}} = \frac{47.5}{3.6}\,\hat{\imath}\,\text{m/s}^2 \approx +13.2\,\hat{\imath}\,\text{m/s}^2. \tag{4.4.4}$$

The cheetah's acceleration is about $13.2\ \text{m/s}^2$ eastward.

(b) Because the truck moves at *constant* velocity, equation (4.17) applies: the camera operator measures the same acceleration as the crew member. So $\mathbf{a}_{\text{camera}} = +13.2\ \hat{\imath}\ \text{m/s}^2$ as well: magnitude $13.2\ \text{m/s}^2$, direction eastward.

Check. Compute it directly in the truck's frame. Initially the cheetah runs $30\ \text{km/hr}$ faster than the truck, westward, so $v_{CT} = -30\ \text{km/hr}$. Finally, by equation (4.16), $v_{CT} = v_{CG} - v_{TG} = 45 - (-20) = +65\ \text{km/hr}$. The change is $65 - (-30) = 95\ \text{km/hr}$ in $2.0\ \text{s}$, exactly the same as in the ground frame. The two observers disagree about the velocities but agree about how fast they change.

Checkpoint 4.4

A passenger walks forward at $1.5\ \text{m/s}$ down the aisle of a train moving forward at $20\ \text{m/s}$ relative to the ground. A second train, on a parallel track, moves backward (in the opposite direction) at $18\ \text{m/s}$ relative to the ground. How fast does the passenger appear to move, as measured by a passenger sitting still on the second train? In what direction?

Answer

Answers to Checkpoints

Checkpoint 4.4. Relative to the ground the passenger moves at $v_{PG} = 20 + 1.5 = 21.5\ \text{m/s}$ forward. Take forward (the first train's direction) as positive, so the second train moves at $v_{TG} = -18\ \text{m/s}$. By equation (4.16), $v_{PT} = v_{PG} - v_{TG} = 21.5 - (-18) = 39.5\ \text{m/s}$. The seated passenger on the second train sees the walking passenger move at $39.5\ \text{m/s}$ in the first train's direction of travel: when two things approach head-on, their relative speed is the *sum* of their speeds.

This page titled 4.4: Relative Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.S: Motion in Two Dimensions (Summary)

Summary

- **Position** in two dimensions is a vector $\mathbf{r}(t) = x\hat{\imath} + y\hat{\jmath}$; displacement is $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$.
- **Velocity** and **acceleration** are the first and second time derivatives of the position vector: $\mathbf{v} = d\mathbf{r}/dt$, $\mathbf{a} = d\mathbf{v}/dt$.
- In two- or three-dimensional motion, the *components along each axis are independent*. With constant acceleration, the kinematic equations of Chapter 2 apply separately along x, y, and z, linked only by the shared time variable.
- **Projectile motion** is two-dimensional motion under gravity alone: $a_x = 0$ and, with $+y$ up, $a_y = -g$. Horizontal motion is uniform; vertical motion is free fall. The acceleration is the same constant downward vector throughout the flight, including at the highest point.
- Eliminating t between the two columns gives the **trajectory** $y = x\tan\theta - (gx^2/2v_0^2)\sec^2\theta$, a downward-opening parabola. For a launch and landing at the *same height*, the range is $R = v_0^2 \sin 2\theta/g$, which is greatest at 45° and takes equal values at complementary angles.
- In **uniform circular motion**, the speed is constant but the direction continuously changes. The centripetal acceleration $a_c = v^2/r$ points toward the center of the circle; because v is squared, doubling the speed quadruples the acceleration. The period of one full revolution is $T = 2\pi r/v$, and the frequency is $f = 1/T$.
- Different observers measure different **relative velocities** for the same object: $\mathbf{v}_{PA} = \mathbf{v}_{PB} + \mathbf{v}_{BA}$. Observers in uniform relative motion measure the same acceleration.

Key Equations

Velocity (2D/3D)

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = v_x\hat{\imath} + v_y\hat{\jmath} \tag{4.S.1}$$

Acceleration (2D/3D)

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = a_x\hat{\imath} + a_y\hat{\jmath} \tag{4.S.2}$$

Kinematic equations for 2D constant acceleration

$$\begin{aligned} v_x &= v_{0x} + a_x\,t, & v_y &= v_{0y} + a_y\,t, \\ \Delta x &= v_{0x}\,t + \tfrac{1}{2}a_x t^2, & \Delta y &= v_{0y}\,t + \tfrac{1}{2}a_y t^2, \\ v_x^2 &= v_{0x}^2 + 2a_x\,\Delta x, & v_y^2 &= v_{0y}^2 + 2a_y\,\Delta y, \\ \Delta x &= \tfrac{1}{2}(v_{0x} + v_x)t, & \Delta y &= \tfrac{1}{2}(v_{0y} + v_y)t \end{aligned} \tag{4.S.3}$$

Trajectory of a projectile

$$y = x\tan\theta - \frac{g\,x^2}{2v_0^2}\sec^2\theta \tag{4.S.4}$$

Range on level ground

$$R = \frac{v_0^2 \sin 2\theta}{g} \tag{4.S.5}$$

Centripetal acceleration

$$a_c = \frac{v^2}{r} \tag{4.S.6}$$

Period and frequency of uniform circular motion

$$T = \frac{2\pi r}{v}, \qquad f = \frac{1}{T} \tag{4.S.7}$$

Relative velocity (Galilean transformation)

$$\mathbf{v}_{PA} = \mathbf{v}_{PB} + \mathbf{v}_{BA} \tag{4.S.8}$$

This page titled 4.S: Motion in Two Dimensions (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

4.E: Motion in Two Dimensions (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 4.1 — Kinematics in Two Dimensions

4.1 ★ A particle's position is given by $\mathbf{r}(t) = (3t^2 - 2t)\ \hat{\imath} + (4t + 1)\ \hat{\jmath}$, with $\mathbf{r}$ in meters and t in seconds.

a. Find $\mathbf{v}(t)$ and $\mathbf{a}(t)$.
b. Evaluate the velocity and acceleration at $t = 1.0\ \text{s}$.
c. What is the magnitude of the velocity at $t = 1.0\ \text{s}$?

4.2 ★★ A ball is initially at the origin with velocity $\mathbf{v}_0 = \langle 5.0, 12.0 \rangle\ \text{m/s}$. It experiences constant acceleration $\mathbf{a} = \langle 2.0, -3.0 \rangle\ \text{m/s}^2$.

a. Write expressions for $v_x(t)$ and $v_y(t)$.
b. Write expressions for $x(t)$ and $y(t)$.
c. At what time is the velocity perpendicular to the initial velocity? *(Hint: use the dot product.)*
d. What is the ball's position at that time?

4.3 ★ *[Conceptual.]* Explain what it means to say that "the horizontal and vertical components of motion are independent." In a two-dimensional motion, is it possible for the x-component of velocity to be zero at an instant when the y-component is nonzero? Give a concrete example.

Section 4.2 — Projectile Motion

4.4 ★ A ball rolls horizontally off the edge of a table $1.20\ \text{m}$ high. It lands $1.50\ \text{m}$ from the base of the table.

a. How long is the ball in the air?
b. What was the ball's initial speed?
c. What is the ball's speed (not just v_x) just before it hits the floor?

4.5 ★★ A projectile is launched from ground level with initial speed $30\ \text{m/s}$ at 40° above the horizontal.

a. Find the maximum height it reaches.
b. How long is it in the air before returning to ground level?
c. How far horizontally does it travel before landing (the *range*)?
d. Repeat for a launch angle of 50°. Compare the ranges. *(They should be equal. Use equation (4.12) to explain why.)*

4.6 ★★ *[Real-world: sports.]* A soccer player kicks a ball from ground level with an initial speed of $22\ \text{m/s}$ at 32° above the horizontal. A defender stands $40\ \text{m}$ away, with arms outstretched to a height of $2.3\ \text{m}$.

a. Find the time at which the ball reaches the horizontal position of the defender.
b. At that time, how high is the ball? Does it pass over the defender's outstretched arms?
c. If the defender is not there, where does the ball land?

4.7 ★★ *[Real-world: engineering.]* Water leaves a horizontal fire hose at a speed of $16\ \text{m/s}$, from a height of $1.5\ \text{m}$. The hose is aimed straight ahead (horizontally).

a. How long does it take a drop of water to reach the ground?
b. How far from the hose does the water land?
c. At what angle below the horizontal does the water strike the ground?

4.8 ★★★ A projectile is launched from the top of a cliff h meters tall, with initial speed v_0 at angle θ above the horizontal. Derive a general expression for:

a. the total time of flight (until it hits the ground at the base of the cliff);
b. the horizontal range from the base of the cliff;
c. the speed of the projectile just before it strikes the ground.

(Express your answers symbolically in terms of h, v_0, θ, and g. Then check by setting $\theta = 0$ and $h = 6\ \text{m}$, $v_0 = 2.26\ \text{m/s}$: the bowling-ball example in the chapter.)

4.9 ★ *[Conceptual.]* On a level field, one bullet is dropped from rest from a height h at the same instant that an identical bullet is fired horizontally from the same height.

a. Which bullet reaches the ground first? Justify your answer in terms of the vertical motion.
b. Does your answer change if the fired bullet has a much larger initial speed? Why or why not?
c. Neglecting air resistance, does the mass of the bullets affect the outcome?

Section 4.3 — Uniform Circular Motion

4.10 ★ A car drives around a circular track of radius $50\ \text{m}$ at a constant speed of $20\ \text{m/s}$.

a. What is the magnitude of the car's centripetal acceleration?
b. How long does it take to complete one lap?
c. How many revolutions per minute does the car make?

4.11 ★★ *[Real-world: biology.]* A laboratory ultracentrifuge spins samples at $60000\ \text{rpm}$. A test tube extends to a radius of $7.5\ \text{cm}$ from the axis of rotation.

a. Convert the rotation rate to revolutions per second, then find the period T.
b. Find the speed of a particle at the outermost point of the tube.
c. Find the centripetal acceleration at that point.
d. Express the result in units of g.

4.12 ★★ *[Real-world: engineering.]* A satellite orbits the Earth in a circular orbit at altitude $400\ \text{km}$ (the altitude of the International Space Station). The radius of the Earth is $6.37 \times 10^6\ \text{m}$, and the satellite's orbital speed is $7.66\ \text{km/s}$.

a. What is the radius of the satellite's orbit, measured from the center of the Earth?
b. Compute the centripetal acceleration of the satellite.
c. Compare your answer to the gravitational acceleration $g = 9.8\ \text{m/s}^2$ at Earth's surface, and comment on the comparison.
d. How long does it take the satellite to complete one orbit? Express in minutes.

4.13 ★★★ A particle moves in a circle of radius $2.0\ \text{m}$. At time $t = 0$ it is at position $(2.0, 0)$ m, moving in the $+y$ direction at $4.0\ \text{m/s}$. It moves at constant speed.

a. Write expressions for $x(t)$ and $y(t)$ for this motion.
b. Compute $\mathbf{v}(t)$ and verify that $|\mathbf{v}| = 4.0\ \text{m/s}$ at all times.
c. Compute $\mathbf{a}(t)$ and verify that its magnitude is v^2/r and that it always points toward the origin.
d. Find the time at which the particle first returns to its starting position.

4.14 ★ *[Conceptual.]* A car moves around a circular track at a perfectly constant speed.

a. Is the car accelerating? Explain how an object can accelerate while moving at constant speed.
b. In what direction does the acceleration point?
c. If the car instead doubles its (still constant) speed around the same track, by what factor does the centripetal acceleration change?

Section 4.4 — Relative Motion

4.15 ★ You are driving north at $25\ \text{m/s}$. A second car is approaching you, traveling south at $30\ \text{m/s}$ in the opposite lane.

a. What is the velocity of the second car relative to you?

b. What is your velocity relative to the second car?

4.16 ★★ *[Real-world: aviation.]* An airplane has an airspeed (speed relative to the surrounding air) of $200\ \text{km/hr}$. There is a wind blowing from west to east at $50\ \text{km/hr}$. The pilot wants to fly due north (relative to the ground).

a. In what direction should the pilot point the airplane? Express the heading as an angle west of north.
b. What is the plane's ground speed (speed relative to the ground)?
c. If the destination is $600\ \text{km}$ due north, how long does the flight take?

4.17 ★★ *[Real-world: navigation.]* A swimmer can swim at $1.2\ \text{m/s}$ relative to still water. She wants to cross a river that is $80\ \text{m}$ wide and flows uniformly at $0.50\ \text{m/s}$.

a. If the swimmer points herself directly across the river (perpendicular to the current), how long does it take her to reach the far bank?
b. How far downstream does she drift in that time?
c. In what direction (relative to straight-across) should she swim if she wants to land directly opposite her starting point? How long does that crossing take?

4.18 ★★★ *[Sketching and conceptual.]* Two observers, A on a station platform and B on a train passing through the station at constant velocity $\mathbf{v}_{BA} = +20\ \hat{\imath}\ \text{m/s}$, watch a ball thrown straight up by a passenger inside the train. The ball reaches a maximum height of $2.0\ \text{m}$ above its release point and is in the air for $1.28\ \text{s}$.

a. In observer B's frame (inside the train), sketch the trajectory of the ball. What shape is it?
b. In observer A's frame (on the platform), sketch the trajectory. What shape is it?
c. Compute the initial vertical and horizontal components of the ball's velocity in B's frame and in A's frame.
d. Compute the acceleration of the ball in each frame. Do they agree? Why or why not?
e. In which frame does the ball travel a greater *distance*? In which frame is the ball's *displacement* larger?

4.19 ★ *[Conceptual.]* A passenger on a train moving at constant velocity drops a coin. To the passenger, the coin falls straight down.

a. Describe the path of the coin as seen by an observer standing still on the ground outside the train.
b. Do the two observers agree on the time the coin takes to hit the floor? Do they agree on its acceleration?
c. Now suppose the train is speeding up while the coin falls. Does the coin still land at the passenger's feet? Explain.

4.20 ★★ *[Conceptual.]* Two cars travel along a straight highway in the same direction, car 1 ahead of car 2.

a. If both cars travel at the same constant speed, what is the velocity of car 1 relative to car 2? What does the gap between them do over time?
b. If car 2 is faster, describe qualitatively what the relative velocity tells you about whether and how the gap changes.
c. Explain why, when overtaking a car at a similar speed on the highway, the maneuver seems to take a surprisingly long time. Frame your answer in terms of relative velocity.

Challenge Problems

4.21 ★★★ *[Real-world: engineering.]* A firefighter aims a nozzle that ejects water at $25\ \text{m/s}$. A window is a horizontal distance $20\ \text{m}$ away, at a height of $12\ \text{m}$ above the nozzle. Use $g = 9.8\ \text{m/s}^2$.

a. Starting from the trajectory equation $y = x\tan\theta - \dfrac{gx^2}{2v_0^2}\sec^2\theta$, show that the launch angle needed to hit the window satisfies a quadratic in $\tan\theta$, and write it out.
b. Solve for the two launch angles that both place the water in the window.
c. Explain physically why two different launch angles can reach the same point.
d. If the nozzle speed could be adjusted, what is the *minimum* speed that can still reach the window, and at what single angle is it launched?

4.22 ★★★ *[Real-world: navigation.]* At $t = 0$, ship A is at the origin steaming due north at $20\ \text{km/hr}$. At the same instant ship B is $10\ \text{km}$ due east of A, moving at $15\ \text{km/hr}$ in a direction 30° west of north. Take east as $+x$ and north as $+y$.

a. Find the velocity of B relative to A (components and magnitude).
b. Write the position of B relative to A as a function of time.
c. At what time are the two ships closest together?
d. What is their minimum separation? Do they collide?

4.23 ★★★ *[Real-world: sports.]* An athlete whirls a stone on the end of a $0.90\ \mathrm{m}$ cord in a *horizontal* circle at constant speed, completing one revolution every $0.40\ \mathrm{s}$. The circle is $1.5\ \mathrm{m}$ above level ground. At the instant the stone is moving due north, the cord breaks.

a. Find the speed of the stone while it is still on the cord.
b. Find its centripetal acceleration, and express the result in units of g.
c. Once the cord breaks, the stone is a projectile. How long does it take to reach the ground?
d. How far does it travel horizontally between the break and the landing?
e. Does the stone fly off due north, or outward away from the athlete? Justify your answer, then use it to find how far the landing point is from the athlete, that is, from the center of the circle. *(Your answer to this part should differ from your answer to part (d). Explain why.)*

This page titled 4.E: Motion in Two Dimensions (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5: THE LAWS OF MOTION I

In Chapters 2 through 4 we learned to *describe* motion: to say where an object is, how fast it is moving, and how its velocity is changing. We never once asked *why*. Why does a ball thrown upward slow, stop, and fall back? Why does a hockey puck slide across the ice almost forever, while the same puck on pavement stops in a meter? Kinematics is silent on these questions. The answers belong to **dynamics**, the study of the causes of motion, and the cause has a name: **force**.

For two thousand years the prevailing view was Aristotle's: that a steadily applied push or pull is needed to keep anything moving, and that, left alone, every object seeks its natural state of rest. This matches everyday experience (stop pushing the cart and it stops), but it is wrong. Galileo saw why. Rolling balls down one ramp and up another, he noticed that the smoother he made the surfaces, the closer the ball came to rising back to its original height, and the farther it rolled on the level. He reasoned that on a *perfectly* smooth, level surface a moving object would never stop at all. What stops the cart is not the absence of a push but the presence of friction: a force we had simply failed to notice. Isaac Newton gathered these insights into three compact laws that, together with his law of gravitation, governed all of mechanics for the next two centuries and remain the working tools of every scientist and engineer today.

Learning Objectives

- State Newton's three laws of motion and explain the physical meaning of each.
- Relate the net force on an object to its mass and acceleration.
- Draw a free-body diagram that isolates a single object and represents every force acting on it.
- Recognize the common forces of mechanics (weight, the normal force, friction, and tension) and calculate all of them but friction, which is taken up in Chapter 6.
- Apply the condition for equilibrium, $\sum \mathbf{F} = 0$, to objects that are not accelerating.
- Explain why a scale in an accelerating elevator does not read an object's true weight, and compute the *apparent weight* that it does read.
- Solve dynamics problems (inclines, hanging objects, connected bodies, and accelerating systems) using a systematic strategy.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

5.1: Newton's Laws of Motion
5.2: Types of Forces
5.3: Free-Body Diagrams
5.4: Equilibrium
5.5: Applying Newton's Laws
5.S: The Laws of Motion I (Summary)
5.E: The Laws of Motion I (Problems)

This page titled 5: The Laws of Motion I is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.1: Newton's Laws of Motion

A **force** is a push or a pull. It is a vector: it has both a magnitude and a direction, and forces combine by the vector addition of Chapter 3. The SI unit of force is the **newton** (N); we will see in a moment that $1\ \mathrm{N} = 1\ \mathrm{kg \cdot m/s^2}$. When several forces act on an object at once, what matters for its motion is their vector sum, the **net force**

$$\mathbf{F}_{\mathrm{net}} = \sum_i \mathbf{F}_i, \tag{5.1}$$

also written $\sum \mathbf{F}$. Newton's laws are statements about this net force.

The first law: inertia

An object's velocity does not change unless a nonzero net force acts on it.

An object at rest stays at rest, and an object in motion continues in motion at constant velocity (the same speed in the same straight line) so long as the net force on it is zero. This is Galileo's insight made into law, and the tendency it describes is called **inertia**; the first law is often called the *law of inertia*.

The first law overturns the Aristotelian picture completely. Constant velocity, including the special case of being at rest, is not something that requires a force to maintain. It is the natural, force-free state. A force is needed only to *change* a velocity, and the rate at which velocity changes is exactly what we called acceleration. So the first law tells us what a force *does*: a net force produces acceleration. It does not yet tell us how much. That is the job of the second law.

Note: The first law picks out a point of view

Strictly speaking, the first law holds only for an observer who is not accelerating; such an observer is said to occupy an **inertial reference frame**. You can feel the difference. When a bus pulls away from a stop, a bag on the seat beside you slides backward, though nothing pushed it. Seen from the sidewalk nothing strange happened at all: the bag simply kept its velocity while the bus accelerated out from under it. Seen from inside the accelerating bus, the bag appears to break the first law. Everything in this book is worked in an inertial frame, which for our purposes means a frame fixed to the ground.

The second law

The net force on an object equals the product of its mass and its acceleration.

Newton originally stated the law in terms of momentum (the net force equals the time rate of change of an object's momentum) and we will return to that more general and more powerful form in Chapter 9. For an object of constant mass it reduces to the form we will use throughout this chapter:

Key Equation: Newton's second law

$$\mathbf{F}_{\mathrm{net}} = m\,\mathbf{a}. \tag{5.2}$$

Equation (5.2) is a *vector* equation, and that is its great convenience: it holds component by component,

$$\sum F_x = m\,a_x, \qquad \sum F_y = m\,a_y, \qquad \sum F_z = m\,a_z, \tag{5.3}$$

so a single two-dimensional problem becomes two independent one-dimensional problems, exactly the strategy that served us so well for projectiles in Chapter 4.

Three things deserve emphasis. First, it is the *net* force that appears, never an individual force. Second, the acceleration points in the same direction as the net force, and is zero precisely when the net force is zero, which is just the first law restated. Third, the mass m measures *inertia*: for a given net force, a larger mass accelerates less. The newton is defined to make the proportionality

constant equal to one, so that $1\ \mathrm{N}$ is the net force that gives a $1\ \mathrm{kg}$ mass an acceleration of $1\ \mathrm{m/s^2}$.

Example 5.1: Net force from two pushes

Two people push a stalled $1200\ \mathrm{kg}$ car along a level road; neglect friction and rolling resistance. One pushes due east with $280\ \mathrm{N}$; the other pushes due east with $340\ \mathrm{N}$. Find the car's acceleration.

Solution

Both forces point the same way, so the net force is simply their sum,

$$F_{\text{net}} = 280\,\mathrm{N} + 340\,\mathrm{N} = 620\,\mathrm{N}, \tag{5.1.1}$$

directed east. By the second law,

$$a = \frac{F_{\text{net}}}{m} = \frac{620\,\mathrm{N}}{1200\,\mathrm{kg}} = 0.52\,\mathrm{m/s^2}, \tag{5.1.2}$$

also directed east. Notice how the units work: $\mathrm{N}/\ \mathrm{kg} = (\ \mathrm{kg}\cdot\mathrm{m/s^2})/\ \mathrm{kg} =\ \mathrm{m/s^2}$, as it must.

The third law

> *If object A exerts a force on object B, then object B exerts a force on object A that is equal in magnitude and opposite in direction.*

Forces always come in pairs. There is no such thing as an isolated push. When you lean on a wall, the wall pushes back on you just as hard; when the Earth pulls down on you with your weight, you pull up on the Earth with an equal force. We can write the third law as a relation between the two members of such an **action–reaction pair**:

Key Equation: Newton's third law

$$\mathbf{F}_{A\text{ on }B} = -\mathbf{F}_{B\text{ on }A}. \tag{5.4}$$

The single most important fact about an action–reaction pair is that *the two forces act on different objects*. One acts on B, the other on A. Because they act on different objects, they never appear together in a single free-body diagram, and they never cancel. The forces in a third-law pair are always of the same type (both gravitational, or both contact forces, and so on) and always equal in magnitude, no matter how different the two objects are: when a moving truck strikes a mosquito, the mosquito pushes back on the truck exactly as hard as the truck pushes on it. What differs enormously is the *acceleration* each undergoes, because, by the second law, the same force produces a tiny acceleration on the massive truck and an enormous one on the mosquito.

Example 5.2: The horse and the cart

A horse is harnessed to a cart. A skeptical student objects: "By the third law, the cart pulls back on the horse exactly as hard as the horse pulls forward on the cart. The two forces cancel. So the horse can never get the cart moving." Where does the argument go wrong?

Solution

It goes wrong at the word *cancel*. Two forces cancel only when they act on the *same* object, and these two do not: the horse's pull acts on the cart, and the cart's pull acts on the horse.

To learn whether the cart accelerates, look only at the forces on the *cart*. They are the horse's forward pull and the backward friction of the ground on the cart's wheels, and the cart's pull on the horse is nowhere among them, because it does not act on the cart. If the horse pulls harder than the wheels resist, the cart accelerates forward.

To learn whether the *horse* accelerates, look only at the forces on the horse. They are the cart's backward pull and the forward friction of the ground on the horse's hooves. The hooves push backward on the ground; by the third law the ground pushes forward on the horse. If that push exceeds the cart's pull, the horse accelerates forward too.

The moral is worth keeping as a rule: *the two members of an action–reaction pair never appear on the same free-body diagram.* Any third-law argument that seems to prove nothing can ever move has almost certainly put both members of a pair on one diagram.

Checkpoint 5.1

A book rests on a table. The book's weight (the Earth pulling down on the book) and the normal force (the table pushing up on the book) are equal in magnitude and opposite in direction. Are these two forces an action–reaction pair in the sense of Newton's third law? Explain. *(Hint: on what objects do the two members of a third-law pair act?)*

Answer

Answers to Checkpoints

Checkpoint 5.1. No. The two forces act on the same object (the book) and a third-law pair always acts on two *different* objects. They are equal and opposite here because the book is in equilibrium (the second law with $a = 0$), not because of the third law. The third-law partner of the book's weight is the book's upward gravitational pull on the Earth; the partner of the table's normal force on the book is the book's downward push on the table.

This page titled 5.1: Newton's Laws of Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.2: Types of Forces

Almost every mechanics problem in this book is built from just a handful of forces. It is worth meeting them one at a time.

Weight

The **weight** of an object is the gravitational force the Earth exerts on it. We will study gravity in depth in Chapter 13; for now we need only that, near the Earth's surface, the weight points straight down and has magnitude

Key Equation: Weight

$$\mathbf{F}_g = m\,\mathbf{g}, \qquad F_g = mg, \tag{5.5}$$

where $g = 9.8\ \mathrm{m/s^2}$ is the free-fall acceleration. Many texts write the weight as W; we will write it $\mathbf{F}_g$ throughout, both to leave the symbol W free for the work of Chapter 7 and to make plain at a glance that weight is a *force*. The other forces follow the same pattern, an F with a subscript that names the force: $\mathbf{F}_N$ for the normal force, $\mathbf{F}_T$ for tension, and, in Chapter 6, $\mathbf{F}_f$ for friction.

Weight and mass are not the same thing. **Mass** is an intrinsic measure of an object's inertia, the same everywhere in the universe; it is measured in kilograms. **Weight** is a force (the local gravitational pull on that mass) measured in newtons, and it depends on where you are. A $2.0\ \mathrm{kg}$ textbook has a mass of $2.0\ \mathrm{kg}$ on the Earth, on the Moon, and in deep space. Its weight is about $19.6\ \mathrm{N}$ on Earth, roughly $3.2\ \mathrm{N}$ on the Moon, and very nearly zero far from any star or planet.

Everyday units blur the distinction, because the pound is a unit of *force*, not of mass: $1\ \mathrm{lb} = 4.45\ \mathrm{N}$. A "150-pound person" therefore weighs about $670\ \mathrm{N}$ and has a mass of roughly $68\ \mathrm{kg}$. A bathroom scale marked in both pounds and kilograms is quietly doing that conversion for you, using an assumed value of g, which is why it would mislead you on the Moon, where the pounds reading would fall to about 25 while the true mass, 68 kilograms, would not change at all.

The normal force

When an object presses against a surface, the surface pushes back. This contact force is the **normal force** $\mathbf{F}_N$. The word "normal" here means *perpendicular*: the normal force always points perpendicular to the surface, away from it.

Two cautions. First, the normal force is *not always vertical*, and it is *not always equal to the weight*. On a ramp it points perpendicular to the ramp, not straight up, and its magnitude is whatever value is required to prevent the object from sinking through the surface: a value we find by applying the second law perpendicular to the surface, not by assuming $F_N = mg$. On level ground with nothing else pushing down, it happens that $F_N = mg$; on an incline, or when someone also presses down on the object, it does not. Second, the normal force is a passive, responsive force: it is as large as it needs to be, up to the point where the surface breaks.

Friction

Friction is a contact force directed *parallel* to the surface, opposing the relative motion, or the tendency toward relative motion, of the surfaces in contact. It is friction that brings Galileo's rolling ball to rest and that lets your shoes grip the floor when you walk. Because friction deserves a careful, quantitative treatment of its own, we postpone the details (static versus kinetic friction, and the coefficient μ) to Chapter 6. In *this* chapter we will never have to *compute* a friction force: every surface in the worked examples is taken to be frictionless, so that we can concentrate on the structure of Newton's laws without that complication. You should still draw a friction arrow on the free-body diagram of any object sitting on a surface described as rough.

Tension

Tension is the force transmitted along a cable, rope, wire, or string when it is pulled taut by forces acting at its two ends. It is directed along the length of the cable and pulls inward on whatever is attached at each end. We will idealize cords as *massless*

and *inextensible* unless stated otherwise; for such an ideal cord the tension is the same at every point along its length, and it pulls equally hard on the objects at the two ends.

A **pulley** redirects a tension. An ideal pulley (massless and frictionless) changes only the *direction* of the rope passing over it, not the magnitude of the tension: the tension is the same on both sides. This is what allows a single cord over a pulley to connect two objects that move in different directions, as in the Atwood machine of Example 5.6 below.

Checkpoint 5.2

A $5.0\ \text{kg}$ box sits at rest on a horizontal table while you press straight down on the top of the box with your hand. Is the normal force the table exerts on the box equal to the box's weight? Explain why or why not.

Answer

Answers to Checkpoints

Checkpoint 5.2. No. Three forces act on the box: its weight $mg = 49\ \text{N}$ down, your push F_P down, and the normal force F_N up. The box is in equilibrium, so $F_N = mg + F_P$, which is *larger* than the weight. The normal force is whatever the perpendicular equation requires, and here it must support both the weight and your push.

This page titled 5.2: Types of Forces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.3: Free-Body Diagrams

Because Newton's second law involves the *net* force on *one* object, the indispensable first step in nearly every dynamics problem is to isolate a single object and inventory every force acting on it. The drawing that records this inventory is called a **free-body diagram**. Mastering it is the most valuable single skill in this chapter.

To draw a free-body diagram:

1. **Choose the object** (the "body") you will analyze, and mentally cut it free of everything else. Represent it as a dot or a simple box.
2. **Draw every force that acts *on* that body**, and only those forces, as an arrow pointing away from the body in the direction the force acts. Include weight (always present near the Earth), and one contact force for each thing that touches the body: a surface (normal force, friction), a rope (tension), a hand (an applied push or pull).
3. **Do not draw forces the body exerts on other things**, and do not draw "the force of motion": there is no such force. A moving object coasting freely has no forward force on it.
4. **Choose coordinate axes.** For an object on an incline it is almost always best to tilt the axes so that one axis lies along the surface and the other is perpendicular to it; then the normal force lies along an axis and only the weight must be decomposed.

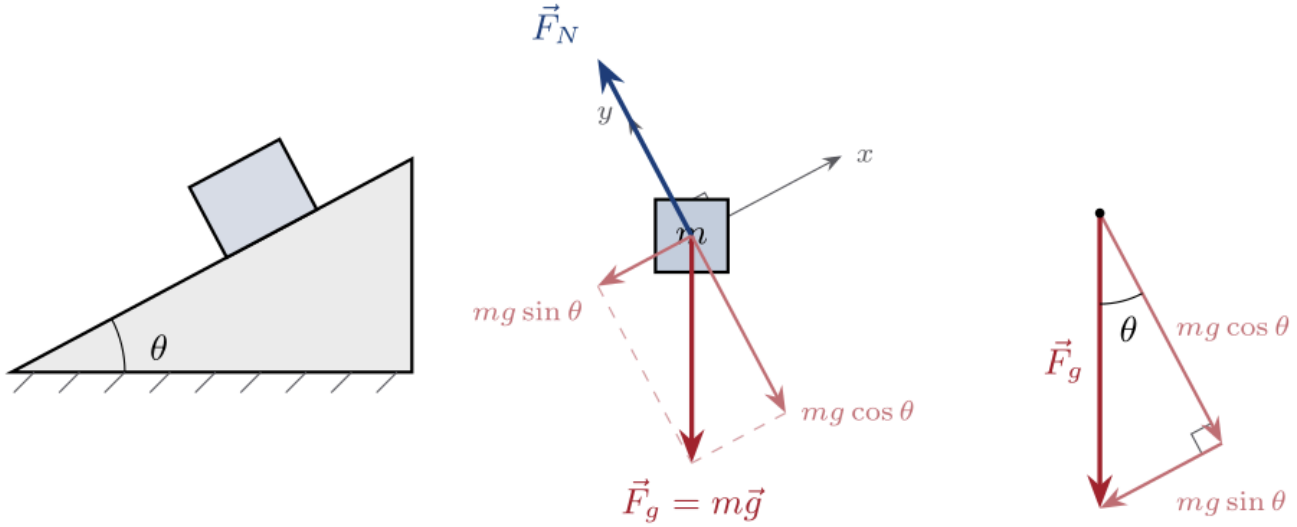


Figure 5.1. Left: a block on a frictionless incline of angle θ. Center: its free-body diagram. Only the weight $\mathbf{F}_g$ and the normal force $\mathbf{F}_N$ act on the block. With the axes tilted along and perpendicular to the surface, only the weight has to be resolved: a component $mg\sin\theta$ down the slope and $mg\cos\theta$ into the surface. Every vector is drawn to the same scale, so notice that $\mathbf{F}_N$ exactly balances $mg\cos\theta$ (the block does not accelerate into the ramp) while nothing at all balances $mg\sin\theta$. That is why the block slides. Right: the weight and its two components redrawn head to tail as a right triangle. The incline angle θ reappears at the top corner, between $\mathbf{F}_g$ and the $mg\cos\theta$ leg, so that leg is adjacent to θ and the $mg\sin\theta$ leg is opposite it.

•Figure description

Left: a block resting on a frictionless incline of angle theta. Center: the block's free-body diagram, showing only two forces, the weight acting straight down and the normal force perpendicular to the incline surface, with the weight's components m g sine theta down the slope and m g cosine theta into the surface. Right: the weight redrawn as a right triangle. The weight F sub g is the hypotenuse, pointing straight down; the component m g cosine theta and the component m g sine theta are the two legs, placed head to tail, with the right angle between them. The angle theta sits at the top corner, between F sub g and the m g cosine theta leg.

Figure 5.1 shows a block resting on a frictionless ramp and the free-body diagram that goes with it. Only two forces act on the block: its weight $\mathbf{F}_g$, straight down, and the normal force $\mathbf{F}_N$, perpendicular to the ramp surface. With axes tilted to the incline, the weight is the only vector that must be broken into components.

With the tilted axes of Figure 5.1, the weight has a component $mg\sin\theta$ pointing down the slope (along $-x$) and a component $mg\cos\theta$ pressing into the surface (along $-y$). The normal force lies entirely along $+y$.

Why does the *slope* component get the sine, when the ramp's angle θ is marked way down at the bottom of the triangle? Because that same angle reappears inside the free-body diagram, as the right-hand triangle in Figure 5.1 shows: the angle between $\mathbf{F}_g$ and the perpendicular to the surface is also θ, since the two angles have mutually perpendicular sides. Once you see that, the components are just the usual right-triangle rules. The side *adjacent* to θ presses into the surface, giving $mg\cos\theta$; the side *opposite* θ runs down the slope, giving $mg\sin\theta$.

If you ever doubt which is which, check the extremes. As $\theta \to 0$ the ramp flattens out, the slope component $mg\sin\theta$ vanishes, and $F_N \to mg$: the object just sits on level ground. As $\theta \to 90^\circ$ the ramp becomes a vertical cliff, $F_N \to 0$, and the entire weight $mg\sin\theta \to mg$ pulls along the "slope," which by then points straight down. Both limits are what they should be, so the assignment of sine and cosine is right. Writing Newton's second law for each axis is then routine, as we will see repeatedly in §5.5.

Checkpoint 5.3

A hockey puck, struck by a stick, slides across frictionless ice at constant velocity. List every force acting on the puck *after* it has left the stick, and sketch its free-body diagram. Is there any force pointing in the direction the puck is moving?

Answer

Answers to Checkpoints

Checkpoint 5.3. Only two forces act: the weight $m\mathbf{g}$, straight down, and the normal force $\mathbf{F}_N$ from the ice, straight up. They balance, since the puck does not accelerate vertically. There is *no* force in the direction of motion: the stick's push ended when contact ended, and on frictionless ice nothing is needed to keep the puck moving. By the first law, zero net force means constant velocity, not rest. An arrow along the motion would be the nonexistent "force of motion."

This page titled 5.3: Free-Body Diagrams is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.4: Equilibrium

An object is in **equilibrium** when the net force on it is zero. By the second law this is exactly the condition that its acceleration is zero:

Key Equation: Condition for (translational) equilibrium

$$\sum \mathbf{F} = 0 \qquad \Longleftrightarrow \qquad \mathbf{a} = 0. \tag{5.6}$$

In two dimensions this single vector condition is two scalar equations, $\sum F_x = 0$ and $\sum F_y = 0$, and these are usually all we need to solve an equilibrium problem.

It is essential to read equation (5.6) carefully: it says the acceleration is zero, *not* that the velocity is zero. An object in equilibrium is not necessarily at rest. A crate sliding across a floor at constant velocity, an airplane cruising straight and level at $900\ \mathrm{km/h}$, and a book sitting on a table are all equally in equilibrium, because in every case the velocity is constant and the net force is zero. “Equilibrium” is a statement about acceleration, not about speed.

Checkpoint 5.4

A skydiver who has reached *terminal velocity* falls straight down at a large but constant speed. Is the skydiver in equilibrium? What does that tell you about the relationship between the downward force of gravity and the upward force of air resistance at that moment?

Answer

Answers to Checkpoints

Checkpoint 5.4. Yes. Terminal velocity is constant velocity, so the acceleration is zero and the net force is zero: equilibrium does not require being at rest. The upward air resistance must therefore exactly equal the skydiver’s weight in magnitude at that moment.

This page titled 5.4: Equilibrium is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.5: Applying Newton's Laws

We now have everything we need to solve problems. Nearly every one yields to the same systematic strategy:

1. **Isolate** the object and draw its free-body diagram (§5.3).
2. **Choose axes.** Align them with the acceleration where possible; tilt them to the incline on a ramp.
3. **Resolve** each force into components along the axes.
4. **Write Newton's second law** for each axis: $\sum F_x = ma_x$, $\sum F_y = ma_y$. For an object in equilibrium along an axis, the right side is zero.
5. **Solve** the resulting equations for the unknowns. For connected objects, write the second law for each body and combine the equations.

Objects on an incline

Example 5.3: Pushing a block up a frictionless ramp

A block weighing $170\ \mathrm{N}$ is pushed up a frictionless ramp inclined at 25° above the horizontal. The push is directed up the slope, and the block accelerates up the slope at $1.3\ \mathrm{m/s^2}$. What is the magnitude of the applied force?

Solution

Step 1: Convert the weight to a mass. The second law needs the mass, but we were given the weight. Since $F_g = mg$, divide both sides by g:

$$m = \frac{F_g}{g} = \frac{170\,\mathrm{N}}{9.8\,\mathrm{m/s^2}} = 17.3\,\mathrm{kg}. \tag{5.5.1}$$

(The units work because $\mathrm{N} = \mathrm{kg}\cdot\mathrm{m/s^2}$, and dividing by $\mathrm{m/s^2}$ leaves kg.)

Step 2: Forces and axes. Three forces act on the block: the weight $\mathbf{F}_g$ (down), the normal force $\mathbf{F}_N$ (perpendicular to the ramp), and the applied push $\mathbf{F}_P$ (up the slope). Tilt the axes so $+x$ points up the slope, as in Figure 5.1. Then $\mathbf{F}_P$ lies entirely along $+x$, the normal force has no x-component at all, and only the weight must be resolved: its component along the slope is $F_g \sin\theta$, pointing *down* the slope, so it enters as $-F_g \sin\theta$.

Step 3: Newton's second law along the slope. The acceleration is up the slope, along $+x$:

$$\sum F_x = ma \quad\Longrightarrow\quad F_P - F_g \sin\theta = ma. \tag{5.5.2}$$

We want F_P, so add $F_g \sin\theta$ to both sides:

$$F_P = ma + F_g \sin\theta = (17.3\,\mathrm{kg})(1.3\,\mathrm{m/s^2}) + (170\,\mathrm{N})\sin 25^\circ. \tag{5.5.3}$$

Each term on the right is a force. The first, $ma = 22.5\ \mathrm{N}$, is what it takes to accelerate the block; the second, $F_g \sin 25^\circ = (170\ \mathrm{N})(0.4226) = 71.8\ \mathrm{N}$, is what it takes just to hold the block against sliding back down. Adding,

$$F_P = 22.5\,\mathrm{N} + 71.8\,\mathrm{N} \approx \boxed{94\,\mathrm{N}}. \tag{5.5.4}$$

Check. If the block moved up at *constant* speed, $a = 0$ and the push would be exactly $F_g \sin 25^\circ = 71.8\ \mathrm{N}$; every newton beyond that goes into accelerating the block. Our answer sits about 30% above the constant-speed value, which is the right size for the modest $1.3\ \mathrm{m/s^2}$ we were given. Note also that F_P is well under the full weight of $170\ \mathrm{N}$: a ramp lets you raise a load with less than its weight, which is the whole point of a ramp.

Example 5.4: A car pulled up an incline by an angled cable

An $1100\ \text{kg}$ car is pulled up a frictionless 15° slope by a cable that exerts a $9000\ \text{N}$ force directed 20° above the slope surface. Find the car's acceleration.

Solution

Step 1: Choose axes and resolve the forces. Tilt the axes to the incline: $+x$ up the slope, $+y$ perpendicular to it and away from the surface. Three forces act on the car.

- The cable tension $\mathbf{F}_T$ makes an angle of 20° with the slope, that is, with the x-axis. The x-axis is adjacent to that angle, so $F_{T,x} = F_T \cos 20^\circ$ and $F_{T,y} = F_T \sin 20^\circ$, both positive.
- The weight $\mathbf{F}_g = m\mathbf{g}$ has components $-mg\sin 15^\circ$ along x (down the slope) and $-mg\cos 15^\circ$ along y (into the slope), exactly as in Figure 5.1.
- The normal force $\mathbf{F}_N$ is entirely along $+y$.

Note that the 20° is measured from the *slope*, not from the horizontal, so it goes straight into the tilted axes with no conversion. That is one more reason to tilt the axes.

Step 2: Newton's second law along the slope. The car does not leave the ramp or sink into it, so $a_y = 0$; the acceleration we seek is a_x. The normal force has no x-component, so it does not appear:

$$\sum F_x = ma_x \quad \Longrightarrow \quad F_T\cos 20^\circ - mg\sin 15^\circ = ma_x. \tag{5.5.5}$$

Dividing both sides by m isolates the acceleration:

$$a_x = \frac{F_T\cos 20^\circ - mg\sin 15^\circ}{m}. \tag{5.5.6}$$

Step 3: Substitute. With $mg = (1100\ \text{kg})(9.8\ \text{m/s}^2) = 10780\ \text{N}$,

$$\begin{aligned} F_T\cos 20^\circ &= (9000\,\text{N})(0.940) = 8457\,\text{N}, \\ mg\sin 15^\circ &= (10780\,\text{N})(0.259) = 2790\,\text{N}, \end{aligned} \tag{5.5.7}$$

so

$$a_x = \frac{8457\,\text{N} - 2790\,\text{N}}{1100\,\text{kg}} = \frac{5667\,\text{N}}{1100\,\text{kg}} \approx \boxed{5.2\,\text{m/s}^2}\ \text{up the slope}. \tag{5.5.8}$$

Check: does the car stay on the ramp? The y-equation tells us. With $a_y = 0$, $\sum F_y = 0$ reads

$$F_N + F_T\sin 20^\circ - mg\cos 15^\circ = 0 \quad \Longrightarrow \quad F_N = mg\cos 15^\circ - F_T\sin 20^\circ. \tag{5.5.9}$$

Numerically $F_N = 10\ 413\ \text{N} - 3078\ \text{N} \approx 7300\ \text{N}$. A surface can only push, never pull, so a normal force must be positive; had this come out negative, the cable would be lifting the car off the ramp and the assumption $a_y = 0$ would be wrong. Notice too that F_N is less than $mg\cos 15^\circ$: the upward tilt of the cable carries part of the load the ramp would otherwise support.

Equilibrium of a hanging object

Example 5.5: Will the traffic light's cables hold?

A traffic light weighing $122\ \text{N}$ hangs from a point where three cables meet, as in Figure 5.2. The light hangs from a vertical cable; that cable is tied at a knot to two others that run up to a support, making angles of 37.0° and 53.0° with the horizontal. The two upper cables will break if the tension in either exceeds $100\ \text{N}$. Does the light hang safely, or does a cable break?

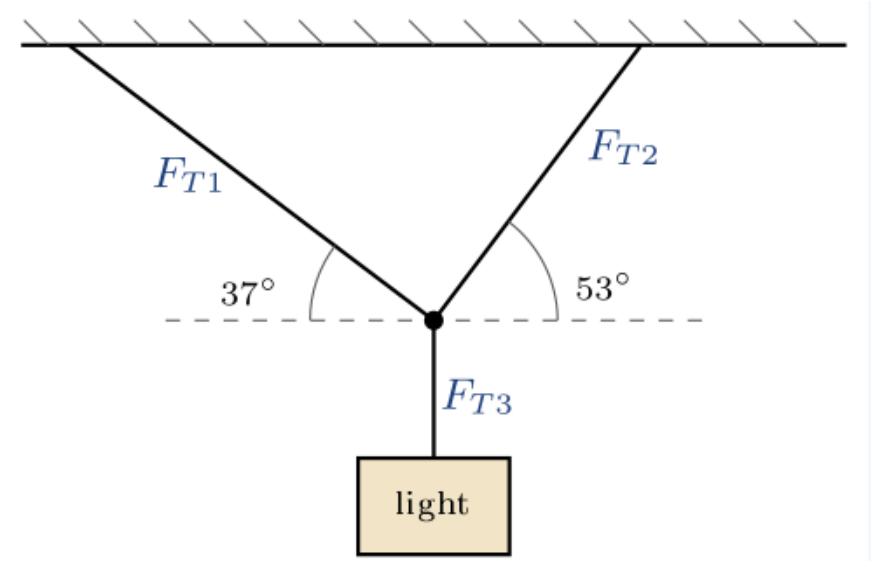


Figure 5.2. A traffic light supported by three cables. The vertical cable carries the full weight; the two upper cables make angles of 37° and 53° with the horizontal.

•Figure description

A traffic light hanging from a junction supported by two upper cables, one rising at 37 degrees to the left and the other at 53 degrees to the right, with tensions labeled F sub T1 and F sub T2, plus a vertical cable holding the light below.

Solution

Step 1: The light. Only two forces act on it: its weight ($122\ \mathrm{N}$, down) and the tension F_{T3} of the vertical cable (up). The light is in equilibrium, so the two must cancel:

$$F_{T3} - F_g = 0 \quad \Longrightarrow \quad F_{T3} = F_g = 122\,\mathrm{N}. \tag{5.5.10}$$

Step 2: The knot. The upper cables do not touch the light, so we move to the knot, where all three cables meet. Three forces act on it: F_{T1} (up and to the left at 37° above the horizontal), F_{T2} (up and to the right at 53°), and the downward pull $F_{T3} = 122\ \mathrm{N}$. Each upper tension makes its angle with the horizontal, so the horizontal component gets the cosine and the vertical component gets the sine; the signs come from the sketch (F_{T1} points left, so its x-component is negative). The knot is in equilibrium, so $\sum F_x = 0$ and $\sum F_y = 0$:

$$\begin{aligned} x: &\quad -F_{T1}\cos 37^\circ + F_{T2}\cos 53^\circ = 0, \\ y: &\quad F_{T1}\sin 37^\circ + F_{T2}\sin 53^\circ - F_{T3} = 0. \end{aligned} \tag{5.5.11}$$

Step 3: Solve two equations in two unknowns. Use the simpler equation to express one unknown in terms of the other, then substitute into the second. The x-equation contains no F_{T3}; add $F_{T1}\cos 37^\circ$ to both sides and divide by $\cos 37^\circ$:

$$F_{T1}\cos 37^\circ = F_{T2}\cos 53^\circ \quad \Longrightarrow \quad F_{T1} = F_{T2}\,\frac{\cos 53^\circ}{\cos 37^\circ} = F_{T2}\,\frac{0.602}{0.799} = 0.754\,F_{T2}. \tag{5.5.12}$$

Now replace F_{T1} in the y-equation with $0.754\ F_{T2}$, and move F_{T3} to the right-hand side. Every remaining term contains F_{T2}, so factor it out:

$$\begin{aligned} (0.754\,F_{T2})\sin 37^\circ + F_{T2}\sin 53^\circ &= 122\,\mathrm{N} \\ (0.454 + 0.799)\,F_{T2} &= 122\,\mathrm{N} \\ 1.252\,F_{T2} &= 122\,\mathrm{N}. \end{aligned} \tag{5.5.13}$$

Divide by 1.252, then return to the x-equation for F_{T1}:

$$F_{T2} = \frac{122\,\mathrm{N}}{1.252} = 97.4\,\mathrm{N}, \qquad F_{T1} = (0.754)(97.4\,\mathrm{N}) = 73.4\,\mathrm{N}. \tag{5.5.14}$$

The larger of the two upper tensions, $F_{T2} = 97.4\ \mathrm{N}$, is just under the $100\ \mathrm{N}$ breaking point, so $\boxed{\text{the light hangs safely}}$, though with

very little margin to spare.

Check. Because $37^\circ + 53^\circ = 90^\circ$, the two upper cables are perpendicular to each other. Tilting the axes to lie along them, each tension must balance the component of the $122\ \text{N}$ weight along its own cable: $F_{T1} = (122\ \text{N})\sin 37^\circ = 73.4\ \text{N}$ and $F_{T2} = (122\ \text{N})\cos 37^\circ = 97.4\ \text{N}$. Same answers, with no simultaneous equations. It also makes sense that the *steeper* cable carries more of the load: it points more nearly straight up, the direction the weight must be opposed.

Connected objects

When two objects are joined by a cord, they share the same magnitude of acceleration (the cord is inextensible), and the cord transmits the same tension to each. Write the second law for each body separately, then solve the equations together.

Example 5.6: The Atwood machine

Two blocks of masses m_1 and m_2 (with $m_1 > m_2$) hang from the two ends of a light cord that passes over an ideal pulley, as in Figure 5.3. This arrangement is called an *Atwood machine*. Find the magnitude of the acceleration of the blocks and the tension in the cord.

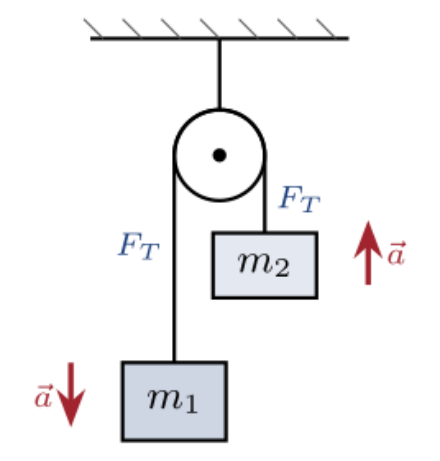


Figure 5.3. An Atwood machine. The heavier block m_1 descends while the lighter block m_2 rises; both have the same acceleration magnitude a, and the cord tension F_T is the same on both sides of the ideal pulley.

•Figure description

An Atwood machine: two blocks of mass m sub 1 and m sub 2 hang from opposite ends of a cord passing over a pulley. Tension F sub T is marked in both strands, and arrows show the blocks accelerating in opposite directions.

Solution

Step 1: Choose a sign convention that matches the motion. Because $m_1 > m_2$, the heavier block descends and the lighter one rises; the cord does not stretch, so both move with the same acceleration magnitude a. Take "downward" as positive for m_1 and "upward" as positive for m_2, so that a single positive a describes both. (Using "up" for both blocks would also work, but then one block would have acceleration $+a$ and the other $-a$, which is an extra sign to keep track of.)

Step 2: Newton's second law for each block. Two forces act on each block: its weight and the cord tension F_T (the same on both sides of an ideal pulley). For m_1, the weight points in the positive (down) direction and the tension in the negative; for m_2 it is the other way around:

$$\begin{aligned} m_1: &\quad m_1 g - F_T = m_1 a, \\ m_2: &\quad F_T - m_2 g = m_2 a. \end{aligned} \tag{5.5.15}$$

Step 3: Eliminate the tension. Two equations, two unknowns (a and F_T). The tension appears as $-F_T$ in the first and $+F_T$ in the second, so *adding* the equations makes it cancel. Add the left sides together and the right sides together:

$$m_1 g - F_T + F_T - m_2 g = m_1 a + m_2 a \implies (m_1 - m_2)g = (m_1 + m_2)a. \tag{5.5.16}$$

The last step factors g out of the left side and a out of the right. Dividing both sides by $(m_1 + m_2)$,

$$\boxed{a = \frac{m_1 - m_2}{m_1 + m_2}\, g}. \tag{5.5.17}$$

Step 4: Find the tension. Solve the second equation for F_T by adding $m_2 g$ to both sides, then substitute for a:

$$F_T = m_2 g + m_2 a = m_2 \left(g + \frac{m_1 - m_2}{m_1 + m_2}\, g \right). \tag{5.5.18}$$

To add the terms in parentheses, write g over the common denominator, as $(m_1 + m_2)g/(m_1 + m_2)$. Then

$$F_T = m_2 g \frac{(m_1 + m_2) + (m_1 - m_2)}{m_1 + m_2} = m_2 g \frac{2m_1}{m_1 + m_2}, \tag{5.5.19}$$

since the m_2's cancel in the numerator. Therefore

$$\boxed{F_T = \frac{2m_1 m_2}{m_1 + m_2}\, g}. \tag{5.5.20}$$

Checks. The result passes two limiting cases. If $m_1 = m_2$, then $a = 0$ and $F_T = 2m_1^2 g/(2m_1) = m_1 g$: equal masses balance, and the cord simply holds each up. And if $m_2 \to 0$, then $a \to g$ and $F_T \to 0$: the lone block m_1 is in free fall. Notice also that F_T lies between the two weights, $m_2 g < F_T < m_1 g$: it must, or the heavier block would not fall and the lighter one would not rise.

As a numerical instance, $m_1 = 3.0\ \text{kg}$ and $m_2 = 2.0\ \text{kg}$ give $a = (1.0/5.0)(9.8) = 1.96\ \text{m/s}^2 \approx 2.0\ \text{m/s}^2$ and $F_T = (2)(3.0)(2.0)(9.8)/5.0 = 23.5\ \text{N}$. As a check, the first equation gives the same tension: $F_T = m_1(g - a) = (3.0)(9.8 - 1.96) = 23.5\ \text{N}$. The tension is indeed between the two weights, $19.6\ \text{N}$ and $29.4\ \text{N}$.

Apparent weight

Example 5.7: A fish on a scale in an elevator

A fish hangs from a spring scale attached to the ceiling of an elevator (Figure 5.4). A spring scale reads the tension in its spring, that is, the force it must exert to support whatever hangs from it. When the elevator is at rest, or moving at any constant velocity, the scale reads the fish's true weight. What does it read when the elevator accelerates? Evaluate the readings for a fish of weight $40.0\ \text{N}$ when the elevator accelerates at $a_y = 2.00\ \text{m/s}^2$, first upward and then downward.

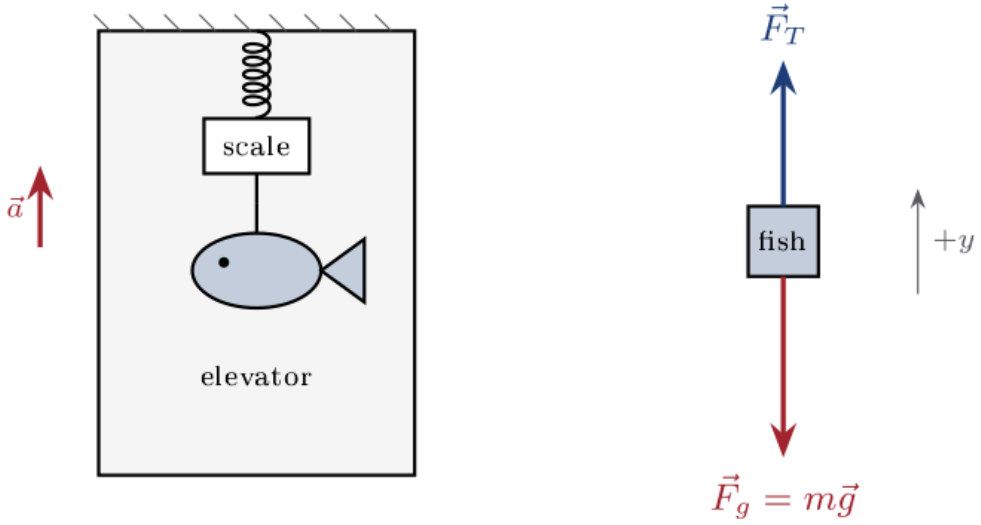


Figure 5.4. Left: a fish hangs from a spring scale in an accelerating elevator. Right: the free-body diagram of the fish. Only two forces act on it, and the scale's reading *is* the tension F_T. Because the fish accelerates with the elevator, F_T and F_g cannot be equal.

•Figure description

A fish hanging from a spring scale inside an elevator. The scale reading is the tension in its support, and an arrow marks the elevator's acceleration.

Solution

Step 1: The general result. Two forces act on the fish: the scale's upward pull F_T (the reading we want) and the weight $F_g = mg$ downward. The fish moves with the elevator, so it has the elevator's acceleration a_y. Take up as positive. Newton's second law:

$$F_T - mg = ma_y. \tag{5.5.21}$$

Add mg to both sides, then factor out the m that both terms share:

$$F_T = mg + ma_y = m(g + a_y). \tag{5.5.22}$$

This is the **apparent weight**: it exceeds the true weight when the elevator accelerates upward ($a_y > 0$) and falls short when the elevator accelerates downward ($a_y < 0$). The sign of a_y does all the work; the formula itself never changes.

Step 2: The mass. The fish's mass is $m = F_g/g = (40.0\ \text{N})/(9.80\ \text{m/s}^2) = 4.08\ \text{kg}$. (The data are given to three significant figures, so we carry g to three here as well.)

Step 3: Evaluate. Accelerating *upward* at $2.00\ \text{m/s}^2$, so $a_y = +2.00\ \text{m/s}^2$:

$$F_T = (4.08\,\text{kg})(9.80 + 2.00)\,\text{m/s}^2 = (4.08)(11.80) \approx \boxed{48.2\,\text{N}}. \tag{5.5.23}$$

Accelerating *downward* at $2.00\ \text{m/s}^2$, so $a_y = -2.00\ \text{m/s}^2$ and the sum in parentheses becomes a difference:

$$F_T = (4.08\,\text{kg})(9.80 + (-2.00))\,\text{m/s}^2 = (4.08)(7.80) \approx \boxed{31.8\,\text{N}}. \tag{5.5.24}$$

Check. Factoring out $mg = F_g$ instead of m gives $F_T = F_g(1 + a_y/g)$, which uses the given weight directly and never needs the mass. Upward: $(40.0\ \text{N})(1 + 2.00/9.80) = (40.0)(1.204) = 48.2\ \text{N}$. Downward: $(40.0)(1 - 0.204) = 31.8\ \text{N}$. The form also shows the size of the effect at a glance: accelerating at about a fifth of g changes the reading by about a fifth.

The scale reads "too heavy" on the way up and "too light" on the way down, exactly the sensation you feel in your own legs as an elevator starts and stops. Were the cable to snap so that $a_y = -g$, the reading would be $m(g - g) = 0$: the fish would be weightless, in free fall along with the scale.

Note: Four pitfalls worth watching for

- **There is no "force of motion."** A coasting object has no forward force on it. If your free-body diagram carries an arrow pointing the way the object happens to be moving, and no rope, hand, or surface is producing it, erase the arrow.
- $F_N = mg$ **is a result, not a rule.** It holds for an object on level ground with nothing else pressing down on it, and almost nowhere else. Find F_N by applying the second law perpendicular to the surface.
- **Action–reaction pairs never share a diagram.** The two forces act on different objects, so they can never cancel each other out.
- **Equilibrium means** $a = 0$**, not** $v = 0$**.** A truck rolling down the highway at a steady $25\ \mathrm{m/s}$ is in equilibrium just as surely as a parked one.

This page titled 5.5: Applying Newton's Laws is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.S: The Laws of Motion I (Summary)

Summary

- A **force** is a push or pull, measured in newtons ($1\ \mathrm{N} = 1\ \mathrm{kg \cdot m/s^2}$). What governs motion is the **net force** $\sum \mathbf{F}$, the vector sum of all forces on an object.
- **Newton's first law** (inertia): with zero net force an object's velocity is constant. Constant velocity, including rest, is the natural force-free state.
- **Newton's second law**: $\mathbf{F}_{\mathrm{net}} = m\mathbf{a}$, applied component by component. The acceleration is parallel to the net force; mass measures inertia.
- **Newton's third law**: $\mathbf{F}_{A\,\mathrm{on}\,B} = -\mathbf{F}_{B\,\mathrm{on}\,A}$. The two forces of an action–reaction pair are equal, opposite, and act on *different* objects, so they never cancel.
- A **free-body diagram** isolates one object and shows every force acting on it. It is the essential first step of every dynamics problem.
- **Equilibrium** means $\sum \mathbf{F} = 0$, equivalently $\mathbf{a} = 0$. It implies constant velocity, not necessarily rest.
- Common forces: **weight** $\mathbf{F}_g = m\mathbf{g}$ (down); the **normal force** $\mathbf{F}_N$ (perpendicular to a surface, not generally equal to the weight); **friction** (parallel to a surface, opposing relative motion; treated in Chapter 6); and **tension** (along a cord, equal at both ends of an ideal massless cord, redirected by an ideal pulley).
- **Mass** (kg, intrinsic inertia) and **weight** (N, the local gravitational force) are different quantities. The pound is a unit of force, not mass.
- A scale measures the force it must supply, not gravity. In an elevator with vertical acceleration a_y, that **apparent weight** is $m(g + a_y)$: larger than mg accelerating upward, smaller accelerating downward, and zero in free fall.

Key Equations

Newton's second law

$$\mathbf{F}_{\mathrm{net}} = m\,\mathbf{a} \tag{5.S.1}$$

Newton's third law

$$\mathbf{F}_{A\,\mathrm{on}\,B} = -\,\mathbf{F}_{B\,\mathrm{on}\,A} \tag{5.S.2}$$

Weight

$$\mathbf{F}_g = m\,\mathbf{g}, \qquad F_g = mg \tag{5.S.3}$$

Condition for (translational) equilibrium

$$\sum \mathbf{F} = 0 \qquad \Longleftrightarrow \qquad \mathbf{a} = 0 \tag{5.S.4}$$

Apparent weight in an accelerating elevator

- Scale reading: $F_T = m(g + a_y)$, with $+y$ up.
- Greater than mg accelerating upward; less than mg accelerating downward; zero in free fall ($a_y = -g$).

This page titled 5.S: The Laws of Motion I (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

5.E: The Laws of Motion I (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 5.1 — Newton's Laws of Motion

5.1 ★ A net force of $36\ \text{N}$ acts on a $4.0\ \text{kg}$ cart on a frictionless track.

a. What is the magnitude of the cart's acceleration?
b. If the cart starts from rest, how fast is it moving after $3.0\ \text{s}$?
c. How far has it traveled in that time?

5.2 ★★ Two horizontal forces act on a $2.5\ \text{kg}$ block resting on a frictionless tabletop: $12\ \text{N}$ directed east and $9.0\ \text{N}$ directed north.

a. Find the magnitude and direction of the net force.
b. Find the magnitude and direction of the block's acceleration.

5.3 ★ *[Conceptual.]* A hockey puck slides across frictionless ice at constant velocity.

a. What is the net force on the puck? How do you know?
b. Aristotle would have said a force is needed to keep the puck moving. In one or two sentences, explain why he was wrong and what actually stops a real puck on real ice.

5.4 ★★ *[Conceptual.]* A heavy truck collides head-on with a small car.

a. During the collision, how does the force the truck exerts on the car compare with the force the car exerts on the truck? Which law tells you this?
b. How do the magnitudes of the two vehicles' accelerations compare during the collision? Explain using Newton's second law.

Section 5.2 — Types of Forces

5.5 ★ An astronaut has a mass of $72\ \text{kg}$.

a. What is her weight on Earth, where $g = 9.8\ \text{m/s}^2$?
b. What is her weight on the Moon, where $g = 1.6\ \text{m/s}^2$?
c. What is her mass on the Moon?

5.6 ★ *[Conceptual.]* A $4.0\ \text{kg}$ book lies on a horizontal table.

a. What is the magnitude of the normal force on the book?
b. You now press down on the book with a force of $20\ \text{N}$. What is the normal force now?
c. Is the normal force always equal to an object's weight? Use parts (a) and (b) to justify your answer.

5.7 ★★ A $12\ \text{kg}$ box rests on a ramp inclined at 20°, held in place by a rope running straight up the slope. Assuming the ramp is frictionless, find (a) the tension in the rope and (b) the normal force on the box.

5.8 ★★ *[Real-world: engineering.]* A $50\ \text{kg}$ crate hangs from a cord that passes over an ideal pulley; the other end of the cord is pulled by a winch. Find the cord tension when the crate is (a) held at rest, (b) lifted at a constant $1.5\ \text{m/s}$, and (c) lifted with an upward acceleration of $0.80\ \text{m/s}^2$.

Section 5.3 — Free-Body Diagrams

5.9 ★ Draw a careful free-body diagram for each of the following objects. Label every force with the object exerting it.

a. A book at rest on a horizontal table.
b. A box being pulled to the right at constant velocity across a rough floor by a rope angled 30° above the horizontal.
c. A ball at the highest point of its flight after being thrown straight up (neglect air resistance).

5.10 ★★ *[Conceptual.]* A student draws a free-body diagram for a ball that has just been thrown horizontally and is now in flight. The diagram shows the ball's weight (down), a normal force (up), and a "force of the throw" (forward).

a. Which of these three forces does not belong, and why?
b. Which force is missing or misdrawn, if any?
c. Redraw the correct free-body diagram in words.

5.11 ★★ A $6.0\ \text{kg}$ block sits on a ramp inclined at 30°. Draw its free-body diagram with axes tilted to the incline, and compute the components of the weight along the slope and perpendicular to the slope.

Section 5.4 — Equilibrium

5.12 ★ A $15\ \text{kg}$ hanging lamp is supported by a single vertical cable.

a. What is the tension in the cable?
b. The lamp is then lowered at a constant speed of $0.40\ \text{m/s}$. What is the tension now? Explain.

5.13 ★★ A $220\ \text{N}$ picture hangs from two wires that run up from the top corners of the frame to a single nail, each making an angle of 40° with the horizontal.

a. Draw the free-body diagram of the nail, where the two wires meet.
b. Find the tension in each wire.
c. If instead the wires made a smaller angle with the horizontal, would the tension increase or decrease? Explain physically.

5.14 ★★ *[Real-world: engineering.]* A $150\ \text{kg}$ sign hangs from the end of a horizontal strut. A support cable runs from the end of the strut up to the wall, making a 55° angle with the strut. Treating the end of the strut as a point in equilibrium under the weight of the sign, the horizontal push of the strut, and the cable tension, find the tension in the support cable.

5.15 ★ *[Conceptual.]* An airplane flies in a perfectly straight, level line at a constant $250\ \text{m/s}$. Four forces act on it: weight, lift, thrust, and drag.

a. Is the plane in equilibrium? How do you know?
b. What can you conclude about lift versus weight, and about thrust versus drag?

Section 5.5 — Applying Newton's Laws

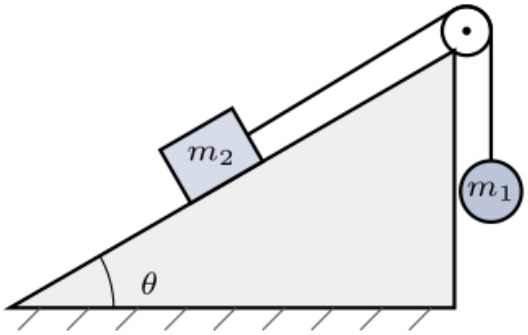
5.16 ★★ A $3.0\ \text{kg}$ block slides down a frictionless ramp inclined at 35°.

a. Draw the free-body diagram and find the block's acceleration along the slope.
b. Find the normal force on the block.
c. If the ramp is $2.5\ \text{m}$ long and the block starts from rest at the top, how fast is it moving at the bottom?

5.17 ★★ *[Real-world: sports.]* A $55\ \text{kg}$ skier starts from rest at the top of a frictionless slope inclined at 12°.

a. Find the skier's acceleration down the slope.
b. How long does it take her to reach a speed of $15\ \text{m/s}$?
c. How far down the slope has she traveled at that point?
d. Look back at your answers. Did the skier's mass enter any of them? Explain why or why not.

5.18 ★★ *[Connected objects: incline and pulley.]* A ball of mass $m_1 = 2.0\ \text{kg}$ hangs from a light cord that passes over an ideal pulley at the top of a frictionless incline of angle $\theta = 30^\circ$; the other end of the cord is attached to a block of mass $m_2 = 3.0\ \text{kg}$ resting on the incline.



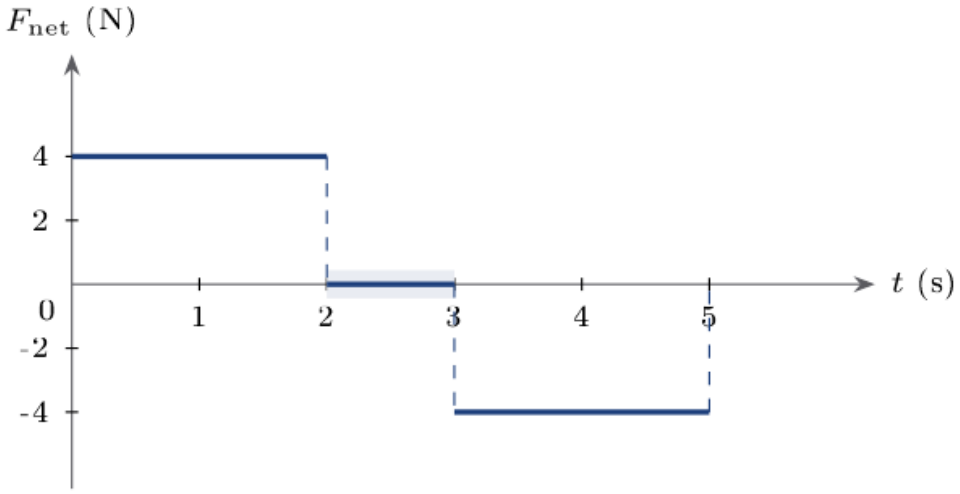
•Figure description
A block of mass m sub 1 hanging from a cord that passes over a pulley at the top of an incline of angle theta, connecting to a second block of mass m sub 2 resting on the incline surface.

a. Draw a free-body diagram for each object.
b. Write Newton's second law for each, and find the magnitude of the acceleration of the system.
c. Find the tension in the cord.
d. Which way does the system accelerate? Justify your sign choice.

5.19 ★★★ *[Two blocks in contact.]* Two blocks of masses $m_1 = 4.0\ \text{kg}$ and $m_2 = 2.0\ \text{kg}$ sit side by side, touching, on a frictionless horizontal surface. A constant horizontal force $F = 18\ \text{N}$ is applied to the outer face of m_1, pushing both blocks.

a. Treating the two blocks as a single system, find the acceleration.
b. Now isolate m_2. Find the contact force that m_1 exerts on m_2.
c. Isolate m_1 instead and use it to find the same contact force. Verify the two results agree, and explain how Newton's third law connects them.

5.20 ★★★ *[Graph-reading.]* A $2.0\ \text{kg}$ cart moves along a straight, frictionless track. The graph below shows the net force on the cart as a function of time.



The cart starts from rest at $t = 0$.

a. Find the cart's acceleration during each of the three intervals.
b. What is the cart's velocity at $t = 2\ \text{s}$? At $t = 3\ \text{s}$? At $t = 5\ \text{s}$?
c. Describe the cart's motion in words over the full $5\ \text{s}$.
d. During the interval from 2 to $3\ \text{s}$, the net force is zero. Is the cart at rest during this interval? Explain.

5.21 ★★ *[Real-world: everyday.]* A $65\ \text{kg}$ student stands on a bathroom scale in an elevator. The scale reads the normal force it exerts on her feet.

a. What does the scale read while the elevator sits at rest?
b. What does it read while the elevator rises at a constant $3.0\ \mathrm{m/s}$?
c. What does it read while the elevator accelerates upward at $1.2\ \mathrm{m/s^2}$?
d. The elevator now has an acceleration of $1.2\ \mathrm{m/s^2}$ directed downward. What does the scale read? Does this necessarily mean the elevator is moving downward? Explain.

Challenge Problems

5.22 ★★★ *[Real-world: engineering.]* A $12\ \mathrm{kg}$ lamp hangs in equilibrium from two cables attached to the ceiling, one making 30° and the other 50° with the horizontal.

a. Draw a free-body diagram of the lamp and write the two equilibrium equations.
b. Find the tension in the 30° cable.
c. Find the tension in the 50° cable.
d. Which cable carries more tension, and why?

5.23 ★★★ Two blocks, $m_1 = 3.0\ \mathrm{kg}$ and $m_2 = 2.0\ \mathrm{kg}$, sit in contact with each other on a frictionless incline of angle 30°. A force $F = 40\ \mathrm{N}$ directed up the incline is applied to the lower block (m_1), pushing both up the slope.

a. Find the acceleration of the two blocks.
b. Find the contact force between them.
c. Suppose the blocks are swapped, so that m_2 is now the lower block, and the same $40\ \mathrm{N}$ up-slope force is applied to m_2. Compute the contact force in this arrangement. Is it the same as in part (b)? Explain the difference physically.

5.24 ★★★ *[Connected objects: table and pulley.]* A block $m_1 = 4.0\ \mathrm{kg}$ rests on a frictionless horizontal table. A light cord runs from m_1 horizontally over a frictionless pulley at the table's edge to a hanging block $m_2 = 2.0\ \mathrm{kg}$.

a. Find the acceleration of the system.
b. Find the tension in the cord.
c. Find the magnitude and direction of the net force the cord exerts on the pulley.

This page titled 5.E: The Laws of Motion I (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6: THE LAWS OF MOTION II

In Chapter 5 we made a quiet but enormous simplification: every surface was frictionless, every cord ideal, every motion in a straight line. Those idealizations let us see the structure of Newton's laws without distraction. Now we put back what we left out. Friction is not a nuisance to be wished away: it is the force that lets you walk, that stops a car, that holds a screw in wood, and (as Galileo realized) the reason a sliding object ever comes to rest. Resistive forces like air drag shape everything from the terminal speed of a skydiver to the design of a race car. And when an object moves in a circle, Newton's second law makes a firm demand: s*omething* must continuously pull it toward the center, or it will fly off in a straight line. This chapter is about these three families of forces, and about applying $\mathbf{F}_{\text{net}} = m\mathbf{a}$ when they are in play.

Learning Objectives

- Distinguish static from kinetic friction, and use the models $F_{f,s} \le \mu_s F_N$ and $F_{f,k} = \mu_k F_N$.
- Solve dynamics problems involving friction on floors, inclines, and walls.
- Describe velocity-dependent resistive (drag) forces and explain the origin of terminal speed.
- Apply Newton's second law to uniform circular motion, identifying the real force that supplies the centripetal force mv^2/r .
- Analyze circular motion in vertical circles, banked curves, conical pendulums, and connected systems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

6.1: Friction
6.2: Resistive Forces
6.3: Uniform Circular Motion
6.S: The Laws of Motion II (Summary)
6.E: The Laws of Motion II (Problems)

This page titled 6: The Laws of Motion II is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6.1: Friction

Begin with three thought experiments. First, send a book sliding across a level floor: it slows and stops. Second, push the book steadily so that it slides at constant velocity: your push is nonzero, yet the velocity is constant, so by the first law the net force is zero; something must be canceling your push. Third, lean on a heavy table that does not move: you push, the table stays put, so again something balances your push exactly. In all three cases the balancing agent is **friction**, a contact force that surfaces exert parallel to their interface, opposing relative sliding (or its tendency).

The three experiments also reveal that friction comes in two regimes. While the surfaces slide across one another, **kinetic friction** acts; before they slide (when friction is busy preventing motion) **static friction** acts.

The friction model

Careful measurements support a simple working model, resting on three approximations: the friction force (for static friction, its *maximum* value) is proportional to the normal force pressing the surfaces together; it is essentially independent of the apparent area of contact; and (for kinetic friction) it is essentially independent of the relative speed. Within this model:

Key Equation: Frictional forces

$$\text{static (no sliding):} \quad F_{f,s} \le \mu_s F_N, \tag{6.1}$$

$$[2pt]\text{kinetic (sliding):} \quad F_{f,k} = \mu_k F_N. \tag{6.2}$$

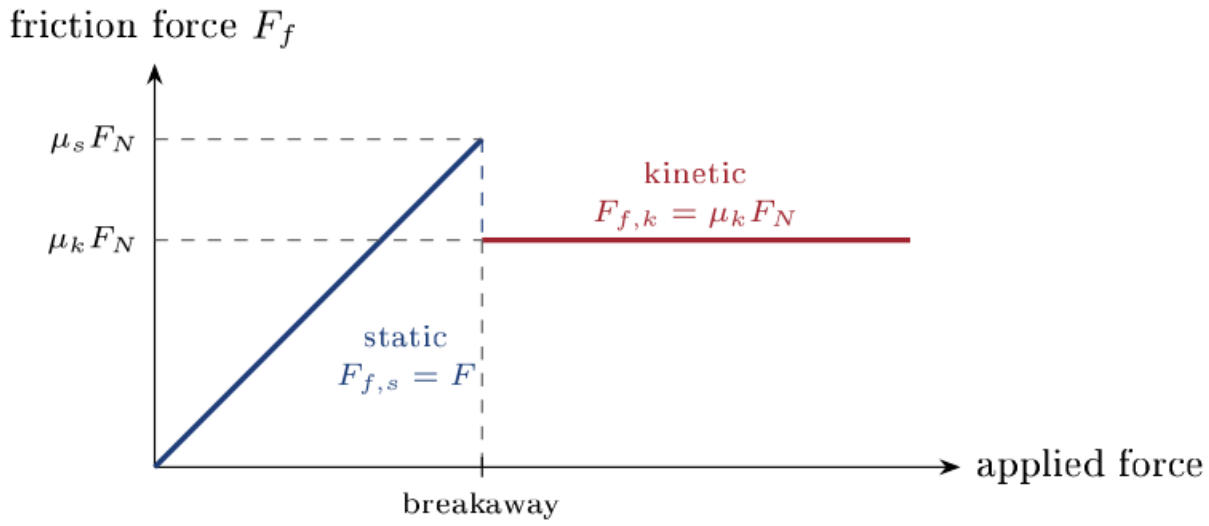


Figure 6.1. Friction force versus applied force for an object initially at rest. In the static region the friction matches the applied force exactly ($F_{f,s} = F$), rising along the diagonal until it reaches its maximum $\mu_s F_N$. At that point the object breaks free; once it slides, the friction drops to the smaller, roughly constant kinetic value $\mu_k F_N$.

•Figure description

A graph of friction force versus applied force. Along the static region the friction rises on the diagonal, exactly matching the applied force, up to a breakaway peak at mu sub s times F sub N. Beyond that the object slides and the friction drops to the constant kinetic value mu sub k times F sub N.

Here F_N is the normal force and μ_s, μ_k are the dimensionless **coefficients of static and kinetic friction**, which depend on the pair of materials. The static law is an *inequality*: static friction takes whatever value (up to its maximum $F_{f,s,\max} = \mu_s F_N$) is needed to prevent sliding. Only at the verge of slipping does $F_{f,s}$ reach $\mu_s F_N$. Once sliding, the friction drops to the constant kinetic value $F_{f,k} = \mu_k F_N$. For almost all surfaces $\mu_s > \mu_k$ (Teflon on Teflon, in Table 6.1, is the near-exception); it takes a larger force to break an object loose than to keep it sliding, which is why a pushed crate often lurches forward once it finally gives.

Figure 6.1 summarizes this behavior. As you slowly increase an applied force on a stationary object, static friction grows to match it (the object stays put), until at $F_{f,s,\max} = \mu_s F_N$ it can no longer keep up; the object breaks free and the friction abruptly drops to the kinetic value.

Representative coefficients appear in Table 6.1. They are approximate, and in a few cases (very clean surfaces) μ_s can exceed 1.

Table 6.1. Approximate coefficients of friction for selected surface pairs.

Surfaces	μ_s	μ_k
Rubber on dry concrete	1.0	0.8
Steel on steel	0.74	0.57
Glass on glass	0.94	0.40
Wood on wood	0.25–0.50	0.20
Metal on metal (lubricated)	0.15	0.06
Ice on ice	0.10	0.03
Teflon on Teflon	0.04	0.04
Synovial joints (human)	0.01	0.003

Working with friction

The procedure is the dynamics strategy of Chapter 5, with friction added to the free-body diagram as a force parallel to the surface. The subtlety is almost always the same: the normal force is whatever the perpendicular equation makes it, and the friction follows from *that* normal force, not from the weight.

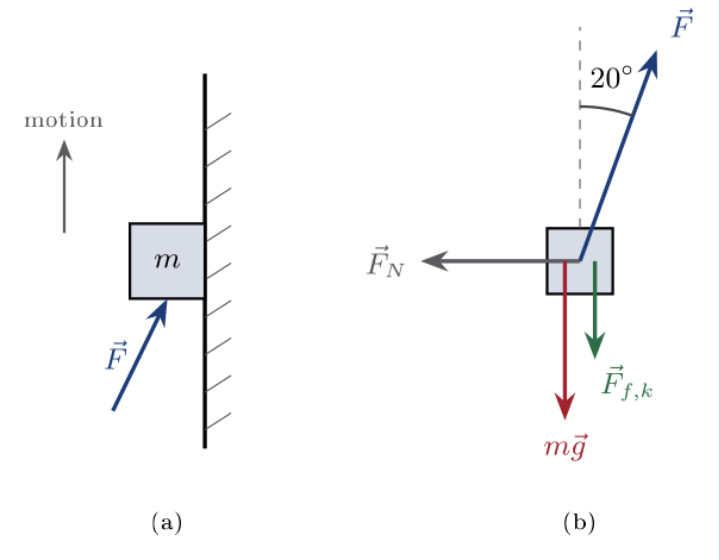
Checkpoint 6.1

A $10\ \text{kg}$ crate rests on a level floor with $\mu_s = 0.50$. You push horizontally with $20\ \text{N}$ and it does not move. What is the friction force on the crate? Is it $\mu_s F_N$?

Answer

Example 6.1: Pushing a crate up a wall

A $10\ \text{kg}$ crate is pushed straight up a vertical wall at constant velocity. The applied force $\mathbf{F}$ makes an angle of 20° with the wall surface (equivalently, 20° from the vertical) and is tilted so that it presses the crate *into* the wall. The coefficient of kinetic friction between crate and wall is $\mu_k = 0.28$. Find the magnitude of $\mathbf{F}$.



•Figure description
A crate held against a vertical wall by a force F pushing into the wall at an angle, with the normal force from the wall, the weight, and the kinetic friction acting along the wall surface.

Solution

Step 1: Axes and components. Take x horizontal (toward the wall positive) and y vertical (up positive). The force $\mathbf{F}$ makes 20° with the *vertical*, so the vertical axis is adjacent to the angle: its components are $F\cos 20^\circ$ up and $F\sin 20^\circ$ into the wall. The crate moves up at constant velocity, so it is in equilibrium ($\sum F_x = 0$ and $\sum F_y = 0$), and kinetic friction acts *downward*, opposing the motion. (In the free-body diagram the weight and the friction are both vertical and downward; they are drawn side by side only so that both remain visible.)

Step 2: The normal force, from the horizontal equation. Horizontally, the wall pushes the crate away ($-F_N$) and the applied force presses it in ($+F\sin 20^\circ$):

$$F\sin 20^\circ - F_N = 0 \quad\Longrightarrow\quad F_N = F\sin 20^\circ. \tag{6.1.1}$$

The normal force is not related to the weight at all here: the wall pushes back only as hard as the crate is pressed into it. The kinetic friction is therefore $F_{f,k} = \mu_k F_N = \mu_k F\sin 20^\circ$.

Step 3: The vertical equation. Up balances down:

$$F\cos 20^\circ = mg + F_{f,k} = mg + \mu_k F\sin 20^\circ. \tag{6.1.2}$$

The unknown F now appears on *both* sides. Gather every term containing F on the left by subtracting $\mu_k F\sin 20^\circ$ from both sides, then factor out F:

$$F\cos 20^\circ - \mu_k F\sin 20^\circ = mg \;\Longrightarrow\; F\left(\cos 20^\circ - \mu_k \sin 20^\circ\right) = mg. \tag{6.1.3}$$

Divide both sides by the quantity in parentheses:

$$F = \frac{mg}{\cos 20^\circ - \mu_k \sin 20^\circ}. \tag{6.1.4}$$

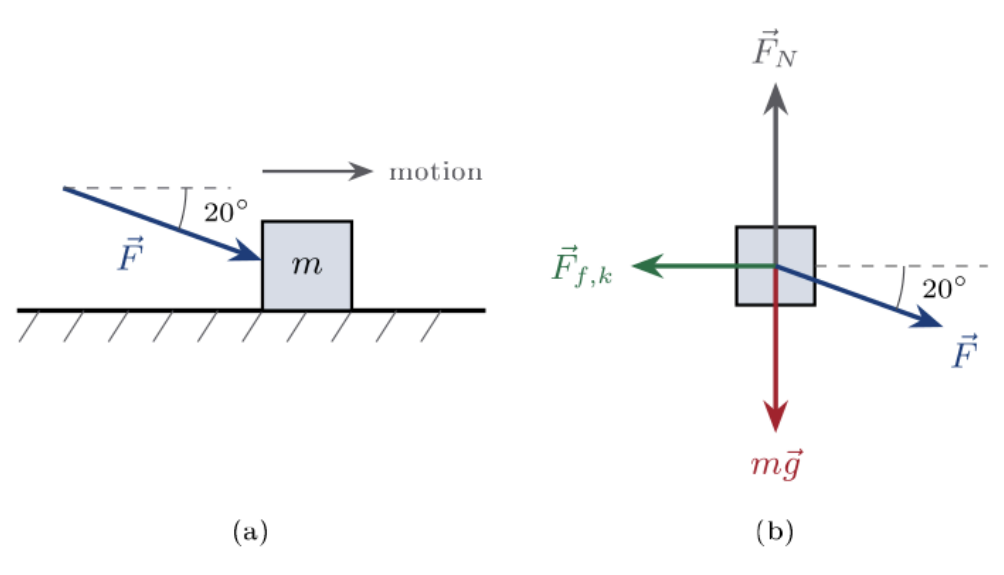
Step 4: Substitute. With $mg = (10\ \text{kg})(9.8\ \text{m/s}^2) = 98\ \text{N}$,

$$F = \frac{98\,\text{N}}{0.940 - (0.28)(0.342)} = \frac{98\,\text{N}}{0.940 - 0.096} = \frac{98\,\text{N}}{0.844} \approx \boxed{116\,\text{N}}. \tag{6.1.5}$$

Check. Without friction the push would only need $F\cos 20^\circ = mg$, or $F = 98/0.940 = 104\ \text{N}$. Friction adds about $12\ \text{N}$, and indeed $F_{f,k} = (0.28)(116)(0.342) = 11\ \text{N}$. The two agree to rounding.

Example 6.2: Pushing a crate along the floor

A $10\ \text{kg}$ crate is pushed across a level floor at constant velocity by a force directed 20° *below* the horizontal. The coefficient of kinetic friction is $\mu_k = 0.10$. Find the magnitude of the pushing force.



•Figure description
Two panels. (a) A crate on a level floor, pushed from behind by a force F directed 20 degrees below the horizontal; an arrow shows the crate moving forward. (b) The crate's free-body diagram: the normal force F sub N straight up, the weight m g straight down, kinetic friction F sub f k pointing backward, and the push F pointing forward and 20 degrees below the horizontal.

Solution

Step 1: Components of the push. Take x horizontal (direction of motion) and y up. The push makes 20° with the horizontal, so it has components $F\cos 20^\circ$ forward and $F\sin 20^\circ$ *downward*. The crate is in equilibrium, and kinetic friction acts backward.

Step 2: The normal force. Vertically, the normal force pushes up, while the weight and the downward part of the push press down:

$$F_N - mg - F\sin 20^\circ = 0 \quad\Longrightarrow\quad F_N = mg + F\sin 20^\circ. \tag{6.1.6}$$

The normal force *exceeds* the weight: note in the diagram that $\mathbf{F}_N$ is drawn longer than $m\mathbf{g}$. It also depends on F, which we do not know yet, so we carry it as an expression.

Step 3: The horizontal equation. Forward push balances backward friction, and friction is μ_k times the normal force of Step 2:

$$F\cos 20^\circ = F_{f,k} = \mu_k F_N = \mu_k\left(mg + F\sin 20^\circ\right). \tag{6.1.7}$$

Distribute μ_k across the parentheses, $\mu_k mg + \mu_k F\sin 20^\circ$, then move the F term to the left and factor:

$$F\cos 20^\circ - \mu_k F\sin 20^\circ = \mu_k mg \;\Longrightarrow\; F\left(\cos 20^\circ - \mu_k \sin 20^\circ\right) = \mu_k mg, \tag{6.1.8}$$

so

$$F = \frac{\mu_k mg}{\cos 20^\circ - \mu_k \sin 20^\circ} = \frac{(0.10)(98\ \text{N})}{0.940 - (0.10)(0.342)} = \frac{9.8\ \text{N}}{0.906} \approx \boxed{11\ \text{N}}. \tag{6.1.9}$$

Check. Put $F = 10.8\ \text{N}$ back in: $F_N = 98 + (10.8)(0.342) = 101.7\ \text{N}$, so $F_{f,k} = (0.10)(101.7) = 10.2\ \text{N}$, while the forward component is $F\cos 20^\circ = (10.8)(0.940) = 10.2\ \text{N}$. They balance.

Pushing *down* on a crate is counterproductive: the downward component increases the normal force, which increases friction. Pulling up at an angle would have the opposite, helpful effect: in the formula, the minus sign in the denominator would become a plus, making the denominator larger and F smaller.

Example 6.3: How far can you lean on a cane?

A person leans on a cane that makes an angle θ with the floor, pushing along its length. The coefficient of static friction between cane tip and floor is $\mu_s = 0.40$. How shallow can θ become before the cane slips?

Solution

Step 1: The forces at the tip. Model the cane as massless, so the force it transmits to the floor is directed along its length, with magnitude F_P. At the tip the floor pushes back along the cane with that same force F_P. Split it into components: the cane makes angle θ with the floor, so the part perpendicular to the floor (opposite θ) is $F_P \sin\theta$ (the normal force F_N) and the part along the floor (adjacent to θ) is $F_P \cos\theta$, which static friction $F_{f,s}$ must supply to keep the tip from sliding out:

$$F_N = F_P \sin\theta, \qquad F_{f,s} = F_P \cos\theta. \tag{6.1.10}$$

Step 2: The no-slip condition. The cane does not slip as long as the friction needed is no more than the friction available, $F_{f,s} \le \mu_s F_N$:

$$F_P \cos\theta \le \mu_s F_P \sin\theta. \tag{6.1.11}$$

Divide both sides by $F_P \sin\theta$. Both F_P and $\sin\theta$ are positive (for $0^\circ < \theta < 90^\circ$), and dividing an inequality by a positive quantity leaves its direction unchanged. The F_P cancels, so how hard you lean does not matter:

$$\frac{\cos\theta}{\sin\theta} \le \mu_s \quad \Longrightarrow \quad \frac{1}{\tan\theta} \le \mu_s, \tag{6.1.12}$$

since $\cos\theta/\sin\theta = 1/\tan\theta$. Now multiply both sides by $\tan\theta$ (positive) and divide both sides by μ_s (positive):

$$\frac{1}{\mu_s} \le \tan\theta, \qquad \text{that is,} \qquad \tan\theta \ge \frac{1}{\mu_s}. \tag{6.1.13}$$

Notice that the inequality "flipped" relative to the one in Step 2, not because of any rule about reciprocals, but simply because we moved $\tan\theta$ and μ_s to opposite sides.

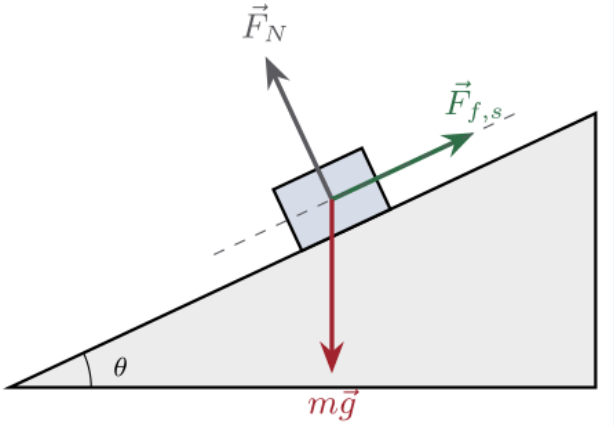
Step 3: The critical angle. The cane is on the verge of slipping when the inequality becomes an equality:

$$\theta_{\min} = \arctan\frac{1}{\mu_s} = \arctan\frac{1}{0.40} = \arctan 2.5 \approx \boxed{68^\circ}. \tag{6.1.14}$$

Since $\tan\theta$ increases with θ, the condition $\tan\theta \ge 2.5$ means $\theta \ge 68^\circ$: for any shallower angle the cane slides out. Equivalently, the cane may tilt at most $90^\circ - 68^\circ \approx 22^\circ$ from vertical. The limits are instructive: as $\mu_s \to 0$ (ice), $1/\mu_s \to \infty$ and $\theta_{\min} \to 90^\circ$: the cane must be held vertical; as μ_s grows large, $1/\mu_s \to 0$ and $\theta_{\min} \to 0$: the cane holds at any angle.

Example 6.4: The angle at which a block lets go

A block sits on a ramp whose angle θ is slowly increased from zero. The coefficients of friction between block and ramp are $\mu_s = 0.35$ and $\mu_k = 0.25$. (a) At what angle does the block begin to slide? (b) What is its acceleration the instant after it lets go?



•Figure description
A block resting on an incline whose angle theta is being increased, with the normal force perpendicular to the surface, the weight downward, and friction acting up along the slope.

Solution

Choose axes *along* the ramp (x, positive down the slope) and *perpendicular* to it (y). This is the one situation where tilted axes pay off: the acceleration, when it comes, will be along the surface. Resolving the weight into these axes gives a component $mg\sin\theta$ down the slope and $mg\cos\theta$ into the surface.

(a) While the block is still at rest, both directions are in equilibrium:

$$y: \quad F_N = mg\cos\theta, \qquad x: \quad F_{f,s} = mg\sin\theta. \tag{6.1.15}$$

Notice again that F_N is *not* mg; it is smaller by the factor $\cos\theta$. The block holds as long as $F_{f,s} \le \mu_s F_N$. Substitute both expressions:

$$mg\sin\theta \le \mu_s\, mg\cos\theta. \tag{6.1.16}$$

Divide both sides by $mg\cos\theta$, which is positive, so the inequality keeps its direction. The mg cancels, and $\sin\theta/\cos\theta = \tan\theta$:

$$\tan\theta \le \mu_s. \tag{6.1.17}$$

The critical angle depends on nothing but the coefficient:

$$\tan\theta_c = \mu_s, \tag{6.3}$$

which for $\mu_s = 0.35$ gives $\theta_c = \arctan 0.35 \approx \boxed{19^\circ}$ (more precisely 19.3°, which we carry into part b). This is worth remembering in both directions: tilting a board until an object slides is a genuine *measurement* of μ_s, requiring nothing but a protractor.

(b) The instant the block breaks free, static friction is replaced by the smaller kinetic friction $F_{f,k} = \mu_k F_N = \mu_k mg\cos\theta$, which no longer balances the down-slope pull. Newton's second law along the slope gives

$$mg\sin\theta - \mu_k mg\cos\theta = ma. \tag{6.1.18}$$

Every term contains m, so divide through by m; then factor g out of the left side:

$$a = g\sin\theta - \mu_k g\cos\theta = g\left(\sin\theta - \mu_k\cos\theta\right). \tag{6.1.19}$$

At $\theta_c = 19.3^\circ$,

$$a = 9.8\left(\sin 19.3^\circ - 0.25\cos 19.3^\circ\right) = 9.8\left(0.330 - 0.236\right) = 9.8\left(0.094\right) \approx \boxed{0.93\ \mathrm{m/s^2}}. \tag{6.1.20}$$

Check. Factor $\cos\theta$ out instead: $a = g\cos\theta\ (\tan\theta - \mu_k)$. At the critical angle $\tan\theta_c = \mu_s$, so $a = g\cos\theta_c\ (\mu_s - \mu_k) = (9.8)(0.944)(0.35 - 0.25) = 0.93\ \mathrm{m/s^2}$. This form makes the physics plain: the acceleration is proportional to the *difference* $\mu_s - \mu_k$. It jumps from

zero to a small positive value the moment the block lets go, precisely the lurch predicted by $\mu_s > \mu_k$. Had the two coefficients been equal, the block would have slid at constant velocity instead.

Checkpoint 6.2

You place a book on a board and slowly tilt the board. The book stays put until the board reaches a certain angle, then begins to slide. Explain how this "tipping-point" angle is related to the coefficient of static friction, and why the book, once moving, tends to keep accelerating rather than sliding at constant speed.

Answer

Answers to Checkpoints

Checkpoint 6.1. The friction force is $20\ \mathrm{N}$, directed opposite your push. The crate is at rest, so the net horizontal force is zero and static friction must exactly balance the $20\ \mathrm{N}$. It is *not* $\mu_s F_N = (0.50)(10)(9.8) = 49\ \mathrm{N}$: that is only the *maximum* static friction can supply. Static friction takes whatever value is needed to prevent sliding, up to that limit. Were it $49\ \mathrm{N}$ while you pushed with $20\ \mathrm{N}$, the crate would accelerate toward you.

Checkpoint 6.2. The book is on the verge of sliding when $\tan\theta = \mu_s$ (equation (6.3)), so the tipping-point angle measures μ_s directly. Once the book moves, static friction is replaced by kinetic friction, which is smaller ($\mu_k < \mu_s$). At that same angle the down-slope pull $mg\sin\theta$ now exceeds the kinetic friction $\mu_k mg\cos\theta$, so there is a net force down the slope and the book accelerates, rather than sliding at constant speed.

This page titled 6.1: Friction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6.2: Resistive Forces

Friction between solid surfaces is, in our model, independent of speed. A fluid (air or water) resists motion very differently: the faster an object moves through it, the harder the fluid pushes back. This speed-dependent force is the **drag force** (or resistive force) $\mathbf{F}_D$, and it always points opposite to the velocity.

At low speeds, or for very small objects, the drag is roughly proportional to speed, $F_D \propto v$. At the everyday speeds of cars, balls, and falling people, drag through air is better described as proportional to the *square* of the speed:

Key Equation: Drag force (high-speed form)

$$F_D = \tfrac{1}{2}\, C\rho A\, v^2, \tag{6.4}$$

where ρ is the fluid density, A is the object's cross-sectional area, and C is a dimensionless **drag coefficient** (around 0.5–1 for ordinary shapes) that captures how streamlined the object is.

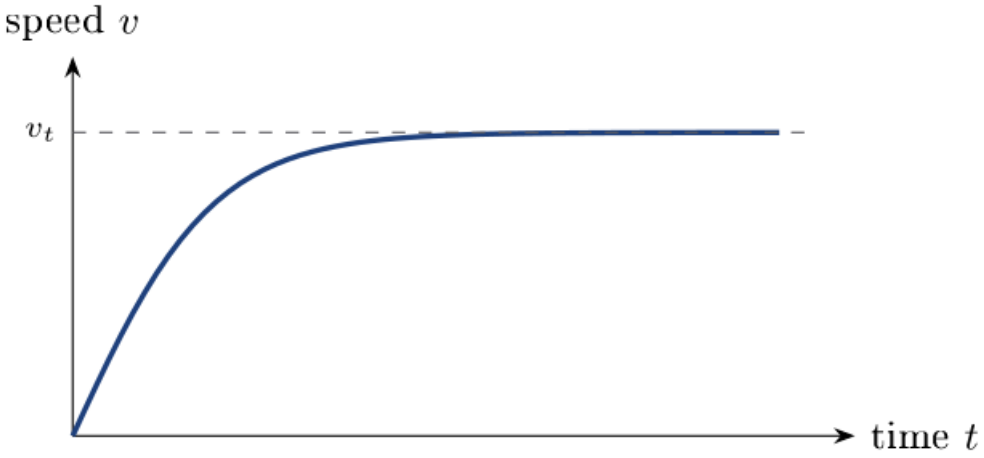


Figure 6.2. An object released from rest in a resistive fluid speeds up, but with ever-decreasing acceleration as drag grows, approaching the terminal speed v_t asymptotically. The curve shown is the quadratic-drag result $v = v_t \tanh(gt/v_t)$ of equation (6.6).

•Figure description

A graph of speed versus time for an object released from rest in a resistive fluid. The curve rises steeply at first, then flattens as drag grows, approaching a horizontal asymptote at the terminal speed v sub t.

The quadratic growth of drag has a famous consequence. Drop an object and it speeds up under gravity; but as v rises, so does the drag, until drag exactly balances weight. From that moment the net force is zero and the object falls at a constant **terminal speed** v_t. Setting $F_D = mg$ in equation (6.4),

$$\tfrac{1}{2}\, C\rho A\, v_t^2 = mg \implies v_t = \sqrt{\frac{2mg}{C\rho A}}. \tag{6.5}$$

Heavier or more compact objects (large m, small A) have higher terminal speeds; this is why a skydiver in a tucked dive falls faster than the same diver spread flat, and why a feather drifts while a pebble plummets. Figure 6.2 shows how the speed of a falling object rises and levels off at v_t.

Notice that equation (6.5) tells us the *final* speed without our ever having to solve for the motion in between. Getting the whole history $v(t)$ takes a little more work, because the acceleration now depends on the speed: $m\ dv/dt = mg - \tfrac{1}{2}C\rho A v^2$. Separating variables and integrating gives

$$v(t) = v_t \tanh\!\left(\frac{gt}{v_t}\right), \tag{6.6}$$

which starts out as $v \approx gt$ (free fall, before drag matters) and flattens toward v_t. A useful rule of thumb follows: the object is within a few percent of terminal speed after a time of roughly $2v_t/g$, about $9\ \mathrm{s}$, and some $250\ \mathrm{m}$ of fall, for a skydiver. If instead the drag is linear in speed, $F_D = bv$, the same procedure gives the exponential approach $v(t) = v_t\left(1 - e^{-bt/m}\right)$ with $v_t = mg/b$.

Example 6.5: Terminal speed of a skydiver

A $75\ \mathrm{kg}$ skydiver falls spread-eagle, presenting a cross-sectional area of about $0.70\ \mathrm{m}^2$ with a drag coefficient $C \approx 1.0$. Taking the density of air as $\rho = 1.2\ \mathrm{kg/m^3}$, estimate the terminal speed.

Solution

Equation (6.5) applies directly. Work out the numerator and denominator under the square root separately before dividing:

$$\begin{aligned} 2mg &= 2(75\,\mathrm{kg})(9.8\,\mathrm{m/s^2}) = 1470\,\mathrm{N}, \\ C\rho A &= (1.0)(1.2\,\mathrm{kg/m^3})(0.70\,\mathrm{m^2}) = 0.84\,\mathrm{kg/m}. \end{aligned} \tag{6.2.1}$$

Then

$$v_t = \sqrt{\frac{2mg}{C\rho A}} = \sqrt{\frac{1470\,\mathrm{N}}{0.84\,\mathrm{kg/m}}} = \sqrt{1750\,\mathrm{m^2/s^2}} \approx \boxed{42\,\mathrm{m/s}}. \tag{6.2.2}$$

The units under the root are $\mathrm{N}/(\ \mathrm{kg/m}) = (\ \mathrm{kg \cdot m/s^2})(\ \mathrm{m/kg}) = \ \mathrm{m^2/s^2}$, and the square root of that is $\ \mathrm{m/s}$: a speed, as it must be. To convert to km/h, multiply by 3.6: $(42)(3.6) \approx 150\ \mathrm{km/h}$, in the right range for a belly-down skydiver (measured speeds are closer to $190\ \mathrm{km/h}$, so the rough values of C and A overestimate the drag somewhat).

Because A sits in the denominator under a square root, $v_t \propto 1/\sqrt{A}$: cutting the area to a quarter doubles the terminal speed. Pulling into a head-down dive cuts A roughly that much and can roughly double this speed.

Checkpoint 6.3

Two balls of the same size and shape are dropped from a great height. One has twice the mass of the other. Which reaches the higher terminal speed, and by what factor? (Assume quadratic drag.)

Answer

Answers to Checkpoints

Checkpoint 6.3. The heavier ball. From equation (6.5), $v_t = \sqrt{2mg/(C\rho A)}$; with C, ρ, and A the same for both, $v_t \propto \sqrt{m}$. Doubling the mass raises the terminal speed by a factor of $\sqrt{2} \approx 1.4$. The heavier ball must fall faster before drag grows large enough to balance its larger weight.

This page titled 6.2: Resistive Forces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6.3: Uniform Circular Motion

In Chapter 4 (§4.3) we found that a particle moving in a circle of radius r at constant speed v has a **centripetal acceleration** of magnitude $a_c = v^2/r$ (equation (4.13)), directed always toward the center of the circle. That was kinematics: a fact about geometry. Newton's second law now turns it into dynamics: an acceleration requires a force.

One kinematic relation from that section will keep coming back. If the object completes one full revolution in a time T (the **period**), it travels a distance $2\pi r$ in that time, so

$$v = \frac{2\pi r}{T}, \qquad T = \frac{1}{f}, \tag{6.7}$$

where f is the number of revolutions per unit time. This is the same period T as in Chapter 4.

Key Equation: Newton's second law for circular motion

$$F_{\text{net}} = ma_c = \frac{mv^2}{r}, \qquad \text{directed toward the center.} \tag{6.8}$$

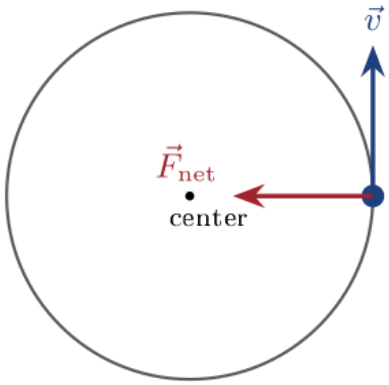


Figure 6.3. In uniform circular motion the velocity is tangent to the circle while the net force points toward the center. This inward force, of magnitude mv^2/r, is supplied by some real force (tension, friction, gravity, or a normal force), never by a separate "centripetal force."

•Figure description

A particle in uniform circular motion, with the velocity vector tangent to the circle and the net force vector pointing inward toward the center.

The net force pointing toward the center is called the **centripetal force** (Figure 6.3). The crucial conceptual point, and the most common source of error, is that the centripetal force is *not a new kind of force*, and it does *not* get added to a free-body diagram. It is the *name we give to the net inward force*, and that net force is always supplied by ordinary forces you already know:

- the **tension** in a string, for a ball whirled on a cord;
- **friction**, for a car rounding a flat curve;
- the **normal force**, for a passenger at the top of a loop or the bank of a curve;
- **gravity**, for a satellite or the Moon.

The recipe, then, is to draw the real forces, point a coordinate axis toward the center, and set the net inward force equal to mv^2/r.

Example 6.6: A ball on a string

A ball of mass $0.65\ \text{kg}$ is attached to a string of length $0.55\ \text{m}$ and swung at a speed of $25\ \text{m/s}$. (a) Find the string tension if the circle is horizontal (treating the string as horizontal, so that the weight is negligible beside the large inward force). (b) Find the maximum tension if the circle is vertical. Assume in both parts that the ball is driven so that its speed stays at $25\ \text{m/s}$ all the way around; a ball swung freely in a vertical circle in fact speeds up on the way down and slows on the way up, which we will be able to handle once we have energy conservation in Chapter 8.

Solution

The centripetal term is

$$\frac{mv^2}{r} = \frac{(0.65\,\text{kg})(25\,\text{m/s})^2}{0.55\,\text{m}} \approx 739\,\text{N}. \tag{6.3.1}$$

(a) For a horizontal circle the tension supplies the entire centripetal force, while the ball's weight ($mg = 6.4\ \text{N}$) is tiny by comparison:

$$F_T = \frac{mv^2}{r} \approx \boxed{7.4 \times 10^2\,\text{N}}. \tag{6.3.2}$$

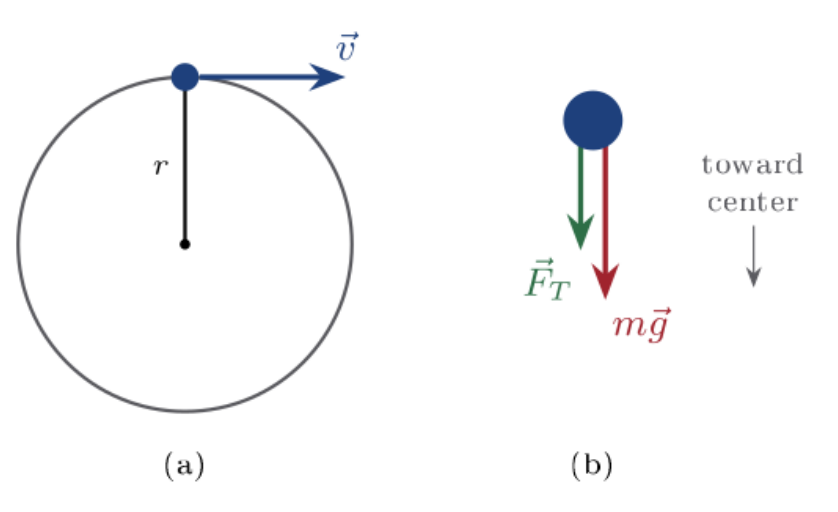
(b) In a vertical circle the tension is largest at the *bottom*, where the string must both supply the centripetal force and support the weight. There the inward (upward) direction gives

$$F_T - mg = \frac{mv^2}{r} \implies F_T = \frac{mv^2}{r} + mg = 739 + 6.4 \approx \boxed{7.5 \times 10^2\,\text{N}}. \tag{6.3.3}$$

At this speed the weight is a small correction; at low speeds it would dominate the difference between top and bottom.

Example 6.7: The slowest you can swing a bucket

A ball is swung in a vertical circle of radius r on a string. How slowly can it pass through the top of the circle and still travel in a circle at all?



•Figure description
Two panels. (a) A ball on a string of length r moving in a vertical circle, shown at the top of the loop with its velocity v pointing horizontally along the circle. (b) The ball's free-body diagram at the top: the string tension F sub T and the weight m g both start at the ball and point straight down, toward the center of the circle.

Solution

At the top of the circle the center is *below* the ball, so "toward the center" means straight down, and both forces on the ball (the

tension and the weight) point that way. Taking downward (toward the center) as positive, the net inward force equals mv^2/r:

$$F_T + mg = \frac{mv^2}{r} \implies F_T = \frac{mv^2}{r} - mg, \tag{6.3.4}$$

where the second form comes from subtracting mg from both sides. Read that result carefully: the required inward force is mv^2/r, gravity supplies mg of it for free, and the string only has to make up the difference. As v drops, less and less tension is needed. When

$$\frac{mv^2}{r} = mg, \tag{6.3.5}$$

the weight alone is exactly the centripetal force the circle requires, and the string goes slack. Below that speed the string cannot help (a string can pull but never push, so F_T can never be negative) and the ball leaves its circular path. Setting $F_T = 0$ therefore gives the minimum speed at the top. Solve $mv^2/r = mg$ for v: divide both sides by m, multiply both sides by r, and take the square root:

$$\frac{v^2}{r} = g \quad \implies \quad v^2 = gr, \tag{6.3.6}$$

so

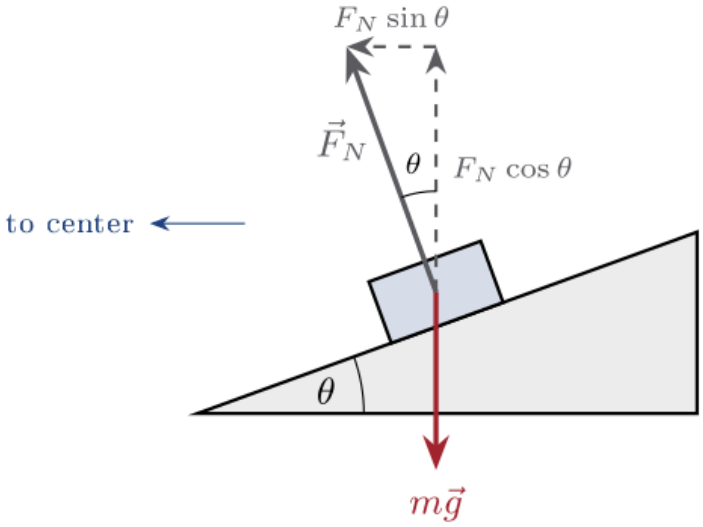
$$v_{\min} = \sqrt{gr}. \tag{6.9}$$

For the $0.55\ \text{m}$ string of the previous example, $v_{\min} = \sqrt{(9.8\,\text{m/s}^2)(0.55\,\text{m})} = \sqrt{5.39\,\text{m}^2/\text{s}^2} \approx \boxed{2.3\,\text{m/s}}$: far below the $25\ \text{m/s}$ used there, which is why that string stayed taut with a great deal to spare.

Two remarks. First, the mass cancels: a bowling ball and a marble on identical strings have the same minimum speed. Second, the same condition governs the water in a bucket swung overhead (the bucket's bottom pushes inward exactly as the string pulls) and a roller-coaster car on the *inside* of a loop, where the normal force F_N replaces F_T. A rigid rod or a car running on a grooved track is the exception: those can push outward as well as pull, so they carry the object around at any speed.

Example 6.8: The ideal speed for a banked curve

A curve of radius $50\ \text{m}$ is banked at 15°. Find the "ideal" speed: the speed at which a car can round the curve with *no* friction required.



•Figure description
A car on a curve banked at angle theta, with the road rising toward the outside of the curve. The normal force F sub N points perpendicular to the banked surface, so it tilts toward the center of the curve; dashed arrows resolve it into a vertical component F sub N cos theta and a horizontal component F sub N sin theta pointing toward the center. The weight m g acts straight down. The angle theta is marked both at the base of the bank and between F sub N and the vertical.

Solution

Step 1: Choose axes. With no friction, only two forces act on the car: gravity (down) and the normal force $\mathbf{F}_N$ (perpendicular to the road). Resolve into horizontal and vertical axes: *not* along the incline this time, because the acceleration is horizontal, toward the center of the curve, and we want it to lie along an axis.

Step 2: Components of $\mathbf{F}_N$. The road is tilted at θ from the horizontal, so its perpendicular, $\mathbf{F}_N$, is tilted at θ from the *vertical*. The vertical axis is therefore adjacent to θ: F_N has a vertical component $F_N \cos\theta$ and a horizontal component $F_N \sin\theta$ pointing toward the center. The vertical components balance (the car neither rises nor sinks), and the horizontal component of $\mathbf{F}_N$ provides the centripetal force:

$$\begin{aligned} \text{vertical:} \quad & F_N \cos\theta = mg, \\ \text{horizontal:} \quad & F_N \sin\theta = \frac{mv^2}{r}. \end{aligned} \tag{6.3.7}$$

Step 3: Divide one equation by the other. Two equations, and F_N is unknown. Rather than solving for F_N first, divide the horizontal equation by the vertical one, left side by left side and right side by right side. (Dividing equal quantities by equal quantities keeps them equal.) On the left, F_N cancels and $\sin\theta/\cos\theta = \tan\theta$; on the right, m cancels:

$$\frac{F_N \sin\theta}{F_N \cos\theta} = \frac{mv^2/r}{mg} \quad \Longrightarrow \quad \tan\theta = \frac{v^2}{rg}. \tag{6.3.8}$$

Step 4: Solve for v. Multiply both sides by rg and take the square root:

$$v = \sqrt{rg\tan\theta} = \sqrt{(50\,\text{m})(9.8\,\text{m/s}^2)(0.268)} = \sqrt{131\,\text{m}^2/\text{s}^2} \approx \boxed{11.5\,\text{m/s}}, \tag{6.3.9}$$

about $(11.5)(3.6) = 41$ km/h. Notice the ideal speed is independent of the car's mass: m canceled in Step 3.

Check. As $\theta \to 0$ (a flat road), $\tan\theta \to 0$ and the ideal speed goes to zero: with no bank and no friction, no speed works except standing still. A steeper bank raises the ideal speed, which is why race tracks are banked more steeply than highway ramps.

Off the ideal speed. Real drivers do not all travel at v_{ideal}, and friction takes up the slack. Go faster and the normal force's inward component is no longer enough, so static friction acts *down* the bank, adding its own inward component; go slower and the car

tends to slide down toward the inside of the curve, so friction acts *up* the bank. Either way the analysis is the one above with a friction term included, and the result is a *band* of safe speeds rather than a single value, which is the whole point of banking a curve. One consequence is worth noting now: if the bank is shallow enough that $\tan\theta \le \mu_s$, there is no minimum speed at all, and a car can sit parked on the curve without sliding. The challenge problems at the end of the chapter work the general case out in full.

Example 6.9: Whirling puck and a hanging weight

A puck of mass m_1 slides in a circle of radius R on a frictionless tabletop. Its string runs through a small hole at the center and down to a hanging object of mass m_2, which stays in equilibrium while the puck circles (Figure 6.4). Find symbolic expressions for (a) the string tension, (b) the puck's speed, and (c) the puck's period of revolution. (d) Describe what happens if m_2 is increased.

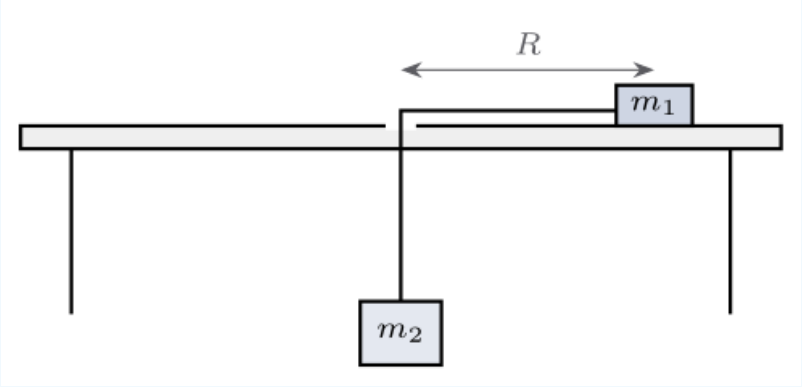


Figure 6.4. A puck of mass m_1 circles on a frictionless table; its string runs through a central hole to a hanging mass m_2 that stays in equilibrium. The tension is the same throughout the string.

•Figure description

A puck of mass m sub 1 moving in a circle of radius R on a horizontal frictionless table, connected by a cord through a hole at the center to a hanging mass m sub 2 below.

Solution

(a) The hanging mass is in equilibrium, so the upward tension balances its weight:

$$F_T - m_2 g = 0 \quad \Longrightarrow \quad F_T = m_2 g. \tag{6.3.10}$$

(b) The only horizontal force on the puck is the string tension, pulling it toward the hole, that is, toward the center of its circle. The whole of the centripetal force is therefore supplied by F_T, and the tension is the same at both ends of the string:

$$m_2 g = \frac{m_1 v^2}{R}. \tag{6.3.11}$$

To isolate v, multiply both sides by R, divide both sides by m_1, and take the square root:

$$v^2 = \frac{m_2 g R}{m_1} \implies v = \sqrt{\frac{m_2 g R}{m_1}}. \tag{6.3.12}$$

Two masses appear here, and it is worth being clear about which does what: m_2 sets the size of the inward pull, while m_1 is the inertia being pulled.

(c) From equation (6.7), $T = 2\pi R/v$. Dividing by v means dividing by a square root, which is the same as multiplying by the square root of the reciprocal fraction:

$$T = \frac{2\pi R}{\sqrt{m_2 g R/m_1}} = 2\pi R\sqrt{\frac{m_1}{m_2 g R}}. \tag{6.3.13}$$

To simplify, move the R outside the root inside it. A factor outside a square root enters it *squared*, $R = \sqrt{R^2}$, and then one R

cancels against the R in the denominator:

$$T = 2\pi\sqrt{\frac{R^2\, m_1}{m_2 g R}} = 2\pi\sqrt{\frac{m_1 R}{m_2 g}}. \tag{6.3.14}$$

Check. Under the root, $\mathrm{kg}\cdot\mathrm{m}/(\,\mathrm{kg}\cdot\mathrm{m/s^2}) = \mathrm{s^2}$, so T comes out in seconds. And the trends are sensible: a heavier hanging mass (m_2 larger) pulls harder and shortens the period; a heavier puck (m_1 larger) is harder to swing around and lengthens it.

(d) Adding load increases m_2 and hence the inward pull, which now exceeds what is needed to hold the puck in its present circle. The puck is drawn inward, and as R shrinks it circles faster and on a tighter path. Strictly, the puck does not simply move to a new steady circle: with nothing to damp it, it overshoots and oscillates in and out while continuing to go around. The formula above applies to a genuine circular path, not to the transition between two of them.

Checkpoint 6.4

A passenger in a car complains of being "thrown outward" against the door as the car rounds a sharp curve. There is in fact no outward force on the passenger. Using Newton's first and second laws, explain what the passenger actually feels and what role the door plays.

Answer

Answers to Checkpoints

Checkpoint 6.4. By the first law, the passenger's body tends to keep moving in a straight line at constant speed. The car, however, curves inward. From the ground's point of view, the passenger is not thrown outward; the car turns underneath the passenger until the door makes contact. The door then pushes *inward* on the passenger, and by the second law that inward push supplies the centripetal force mv^2/r needed to carry the passenger around the curve. What the passenger feels as being "thrown outward" is actually the door pressing inward.

This page titled 6.3: Uniform Circular Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6.S: The Laws of Motion II (Summary)

Summary

- **Friction** acts parallel to a surface, opposing sliding or its tendency. **Static** friction adjusts up to a maximum, $F_{f,s} \le \mu_s F_N$; **kinetic** friction is roughly constant, $F_{f,k} = \mu_k F_N$. Usually $\mu_s > \mu_k$. A block on a ramp lets go when $\tan\theta = \mu_s$, which makes a tilting board a direct measurement of μ_s.
- The normal force is set by the perpendicular equation, not assumed equal to the weight; friction follows from that normal force.
- **Resistive (drag) forces** oppose motion through a fluid and grow with speed, $F_D = \frac{1}{2} C\rho A v^2$ at ordinary speeds. A falling object reaches a constant **terminal speed** $v_t = \sqrt{2mg/(C\rho A)}$ when drag balances weight.
- In **uniform circular motion** the net force points toward the center with magnitude $F_{\text{net}} = mv^2/r$. This **centripetal force** is not a separate force: it is supplied by tension, friction, the normal force, or gravity, and is never added to a free-body diagram.
- Standard cases: a vertical circle (tension greatest at the bottom, and $v_{\min} = \sqrt{gr}$ at the top if the string is to stay taut), a frictionless banked curve ($\tan\theta = v^2/rg$), and connected systems where one body's weight supplies another's centripetal force.

Key Equations

Frictional forces

$$\begin{aligned} \text{static (no sliding):} \quad & F_{f,s} \le \mu_s F_N, \\ \text{kinetic (sliding):} \quad & F_{f,k} = \mu_k F_N. \end{aligned} \tag{6.S.1}$$

Drag force (high-speed form)

$$F_D = \tfrac{1}{2}\, C\rho A\, v^2 \tag{6.S.2}$$

Newton's second law for circular motion

$$F_{\text{net}} = ma_c = \frac{mv^2}{r}, \qquad \text{directed toward the center.} \tag{6.S.3}$$

This page titled 6.S: The Laws of Motion II (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

6.E: The Laws of Motion II (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 6.1 — Friction

6.1 ★ A $25\ \text{kg}$ box rests on a level floor with $\mu_s = 0.45$ and $\mu_k = 0.35$.

a. What is the maximum horizontal force you can apply before the box starts to move?
b. Once it is sliding, what horizontal force keeps it moving at constant velocity?
c. Explain why the answer to (b) is smaller than (a).

6.2 ★★ A $6.0\ \text{kg}$ block is held against a vertical wall by a horizontal push of $40\ \text{N}$. The coefficient of static friction is $\mu_s = 0.50$.

a. Draw the free-body diagram.
b. Does the block slide, or does static friction hold it in place? Justify with a calculation.
c. What is the smallest horizontal push that would hold the block in place?

6.3 ★★ A $12\ \text{kg}$ sled is pulled across level snow ($\mu_k = 0.10$) by a rope at 30° above the horizontal.

a. Find the rope tension needed to pull the sled at constant velocity.
b. Compare the normal force with the sled's weight, and explain the difference.

6.4 ★★ A block is given a shove that sends it sliding *up* a ramp inclined at θ, with coefficient of kinetic friction μ_k.

a. Show that on the way up its acceleration is directed *down* the slope with magnitude $a = g(\sin\theta + \mu_k \cos\theta)$. Explain why the two terms add here but subtract for a block sliding down.
b. Once the block has come to rest, under what condition on μ_s does it stay there?
c. For $\mu_k = 0.20$, find the smallest angle at which a block, once nudged, will continue to slide down at constant velocity.

6.5 ★ *[Conceptual.]* Anti-lock brakes are designed to keep a car's tires from skidding during hard braking.

a. Using the relationship between μ_s and μ_k, explain why a rolling (non-skidding) tire can stop a car in a shorter distance than a skidding one.
b. Why is it harder to steer a car whose wheels are locked and skidding?

6.6 ★★ *[Real-world: biology.]* The synovial joints in your knees have an extraordinarily low coefficient of friction (see Table 6.1). If the cartilage surfaces in a knee press together with a normal force of $500\ \text{N}$ during a stride, estimate the kinetic friction force opposing the joint's motion, and comment on why such low friction matters for a lifetime of use.

6.7 ★★★ *[Experimental design.]* You are given only a stopwatch, a meter stick, and a textbook, and asked to measure the coefficient of kinetic friction between the book and the floor.

a. Describe a procedure (for example, giving the book a push and timing its slide) that yields μ_k from measured distances and times.
b. Derive the equation you would use, starting from Newton's second law with friction.
c. Identify the largest source of experimental uncertainty in your method.

6.8 ★★ A $3.0\ \text{kg}$ block ($\mu_s = 0.40$, $\mu_k = 0.30$) sits on a ramp whose angle is slowly increased from zero.

a. At what angle does the block begin to slide?
b. Immediately after it begins to slide, what is its acceleration down the ramp?

6.9 ★ *[Conceptual.]* Static friction does not always oppose an object's motion: sometimes it is what drives the motion.

a. When you walk forward, in which direction does the static friction from the floor act on your shoe, and how does it propel you?
b. A box rides in the bed of a pickup truck that accelerates forward without the box slipping. In which direction does static friction act on the box, and what is its role?

Section 6.2 — Resistive Forces

6.10 ★ *[Conceptual.]* Two identical sheets of paper are dropped from the same height, but one is crumpled into a tight ball.

a. Which reaches the floor first, and why, in terms of the drag force and terminal speed?
b. In a vacuum, which would reach the floor first?

6.11 ★★ A $4.0\ \text{g}$ ping-pong ball has a terminal speed of about $9.0\ \text{m/s}$ in air.

a. What is the magnitude of the drag force on the ball when it falls at terminal speed?
b. What is the drag force when the ball moves at half its terminal speed (using the quadratic model)?

6.12 ★★ *[Real-world: engineering.]* A spherical raindrop of radius $1.5\ \text{mm}$ falls through air ($\rho = 1.2\ \text{kg/m}^3$, drag coefficient $C \approx 0.5$). Taking the density of water as $1000\ \text{kg/m}^3$, estimate the raindrop's terminal speed. *(Hint: find the drop's mass and cross-sectional area from its radius.)*

Section 6.3 — Uniform Circular Motion

6.13 ★ A $0.20\ \text{kg}$ ball is whirled in a horizontal circle of radius $0.80\ \text{m}$, completing 2.0 revolutions per second.

a. Find the ball's speed.
b. Find the magnitude of the centripetal force, and state what supplies it.

6.14 ★ *[Conceptual.]* A car rounds a flat, unbanked curve at constant speed.

a. Which force provides the centripetal force?
b. What happens, and why, if the road suddenly becomes icy?

6.15 ★★ A $1200\ \text{kg}$ car rounds a flat curve of radius $45\ \text{m}$. The coefficient of static friction between tires and road is $\mu_s = 0.60$.

a. What is the maximum speed at which the car can round the curve without skidding?
b. Does this maximum speed depend on the car's mass? Explain.

6.16 ★★ A pilot of mass $80\ \text{kg}$ flies a loop of radius $250\ \text{m}$ at a constant speed of $60\ \text{m/s}$.

a. Find the force the seat exerts on the pilot at the bottom of the loop.
b. Find the force at the top of the loop.
c. Express each as a multiple of the pilot's weight (the "g-force").

6.17 ★★ *[Conceptual.]* At the top of a vertical loop, a ball on a string and a car on a track both move in a circle.

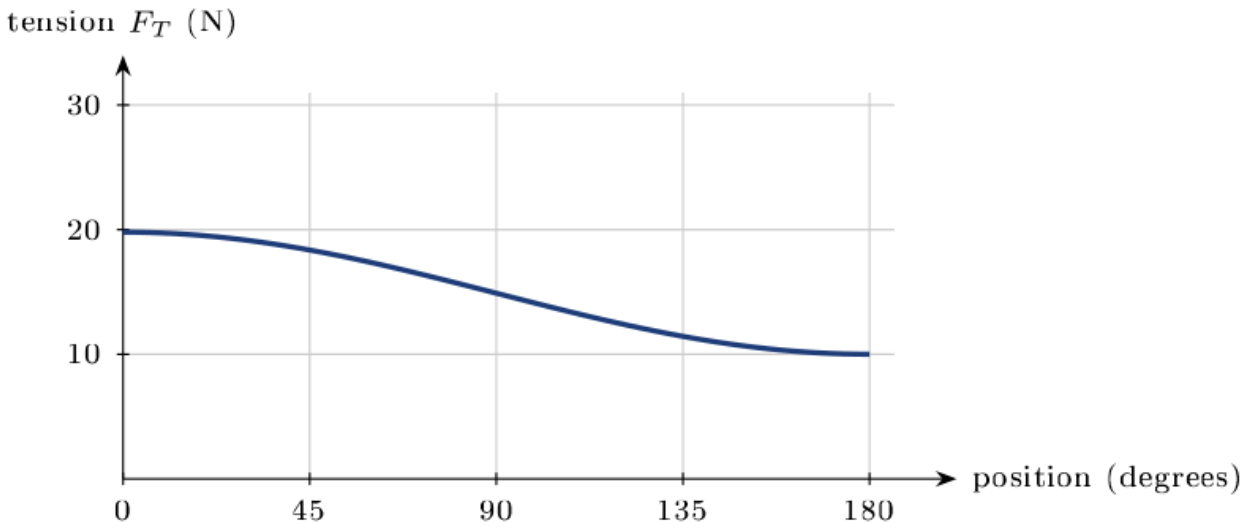
a. For the ball on a string, what is the minimum speed at the top for the string to remain taut?
b. Explain why this minimum speed does not depend on the mass.

6.18 ★★★ *[Real-world: sports / engineering.]* A race car of mass $600\ \text{kg}$ rounds a flat circular turn of radius $100\ \text{m}$. Aerodynamic "downforce" presses the car toward the track, adding to the normal force. The coefficient of static friction between tires and track is 0.75. If the car is on the verge of sliding when its speed is $40.0\ \text{m/s}$, find the magnitude of the downforce, and compare it with the car's weight.

6.19 ★★★ *[Real-world: engineering.]* An amusement-park swing ride hangs seats from chains of length $d = 2.50\ \text{m}$ attached to the rim of a rotating platform of diameter $D = 6.00\ \text{m}$. When the ride turns at constant speed, the chains swing out to $\theta = 30.0^\circ$ from the vertical. The seat plus rider may be treated as a mass on the end of each chain.

a. Show that the radius of a seat's circular path is $r = D/2 + d\sin\theta$, and evaluate it.
b. Find the speed of each seat. (Note that it does not depend on the rider's mass.)
c. If a chain can support at most $1500\ \text{N}$, what is the greatest rider-plus-seat mass that is safe at this angle?

6.20 ★★★ *[Graph-reading.]* A $0.50\ \text{kg}$ ball is swung in a *vertical* circle of radius $1.0\ \text{m}$, driven so that its speed stays essentially constant all the way around. The graph shows the string tension as a function of position, from the bottom ($0°$) up one side to the top ($180°$).



•Figure description
A graph of string tension versus angular position for a ball swung in a vertical circle, from the bottom of the circle at zero degrees up one side to the top at 180 degrees. The tension is greatest at the bottom and falls to its smallest value at the top.

a. At which position is the tension greatest, and at which is it least? Explain physically.
b. Read the tension at the bottom and at the top from the graph, and use each to find the ball's speed there. The two answers should agree: explain why they must.
c. Show that for a ball circling at constant speed the tension at the bottom must exceed the tension at the top by exactly $2mg$, and check this against the graph.

Challenge Problems

6.21 ★★★ A $2.0\ \text{kg}$ book rests on top of a $5.0\ \text{kg}$ board that sits on a frictionless table. Between the book and the board $\mu_s = 0.30$ and $\mu_k = 0.20$. A horizontal force F is applied to the board.

a. What is the largest acceleration the book can have without slipping on the board?
b. What is the largest force F that can be applied to the board while the two still move together?
c. If $F = 30\ \text{N}$ (large enough to cause slipping), find the acceleration of each block.

6.22 ★★★ *[Real-world: engineering.]* A skydiver of total mass $75\ \text{kg}$ falls through air ($\rho = 1.2\ \text{kg/m}^3$, drag coefficient $C = 1.0$). Falling in a tight spread (smaller than the spread-eagle posture worked out in Section 6.2) her body presents a cross-sectional area $A_1 = 0.55\ \text{m}^2$; her open parachute presents $A_2 = 25\ \text{m}^2$. Use $F_D = \tfrac{1}{2}C\rho A v^2$.

a. Find her terminal speed in free fall, before the chute opens.
b. Find her terminal speed under the open parachute.
c. By what factor does opening the chute reduce the terminal speed, and why?
d. At the instant the chute fully opens she is still moving at the free-fall terminal speed. Find her acceleration at that instant, and express it in units of g.
e. Your answer to (d) is far larger than any human could survive. Real canopies take a second or two to inflate. Explain how that changes the calculation, and why the peak deceleration in an actual jump is only a few g.

6.23 ★★★ *[Real-world: engineering.]* A curve of radius $120\ \mathrm{m}$ on a high-speed test track is to be banked. *(Hint for (b) and (c): resolve $\mathbf{F}_N$ and $\mathbf{F}_{f,s}$ into horizontal and vertical components, with $F_{f,s} = \mu_s F_N$ at the limit, and solve the two equations for v. Friction points down the bank in one case and up the bank in the other.)*

a. Find the banking angle for which a car needs no friction to round the curve at the design speed $25\ \mathrm{m/s}$.
b. At that banking angle, with $\mu_s = 0.40$ between tires and road, find the maximum speed at which the car can take the curve without sliding outward.
c. Find the minimum speed below which it tends to slide inward (down the bank).
d. Comment on the range of safe speeds, compared with the single safe speed of an unbanked curve.
e. Public highways are rarely banked more than about 6°, far less than your answer to (a). Given your answers above, suggest why highway engineers accept a bank well below the ideal one.

This page titled 6.E: The Laws of Motion II (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7: WORK AND POWER

Chapters 5 and 6 answered the question "what happens *at this instant*?" Draw the free-body diagram, add the forces, and Newton's second law hands you the acceleration. That is a complete description, but it is not always the most useful one. Very often we care less about the acceleration at each instant than about the cumulative effect of a force acting over a stretch of motion: how much a push across the floor has accomplished, how much a hoist has to supply to raise a load to a rooftop, how quickly an engine can do it.

The quantity that measures this cumulative effect is **work**. A force does work on an object when the object moves while the force acts on it, and the work it does is the amount by which it changes the object's **energy**. We will not try to say here exactly what energy *is*. What forms it takes and how it is stored is the business of Chapter 8. For now it is enough to know that energy is a single number, measured in joules, and that work is how a force increases or decreases it. This chapter builds the tools to calculate work: first for a constant force, then for the most familiar force of all, gravity, and then for forces that change as the object moves. It closes with **power**, which measures how fast the work gets done.

Nothing about the free-body method goes away. Almost every problem in this chapter still begins with a free-body diagram and Newton's laws. They are how we find the forces whose work we need. What is new is the last step: multiplying by a displacement.

Learning Objectives

- Interpret work as a transfer of energy, and the net work on an object as the change in its energy.
- Compute the work done by a constant force, using a free-body diagram and Newton's laws to find the forces involved.
- Compute the work done by gravity, and recognize that it depends only on the change in height.
- Compute the work done by a force that varies with position, including the spring force of Hooke's law.
- Work backward from a work or an energy change to a displacement.
- Compute average and instantaneous power, and use them to answer questions about time.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

7.1: Work Done by a Constant Force
7.2: Work Done by Gravity
7.3: Work Done by a General Force
7.4: Power
7.S: Work and Power (Summary)
7.E: Work and Power (Problems)

This page titled 7: Work and Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.1: Work Done by a Constant Force

Push a stalled car and it starts to roll; you have changed something about it, and the change is bigger the farther you push. Hit the brakes in a moving car and it slows down; friction takes something away from it, and more is taken the farther the car travels while braking. In both cases a force acts while the object moves, and in both cases the object's **energy** changes. The amount of energy a force adds to or removes from an object is the **work** done by that force:

- *positive* work means the force transfers energy *to* the object;
- *negative* work means the force takes energy *from* the object.

Throughout this chapter we treat each object as a particle, as we did in Chapters 5 and 6. Several forces may act on it at once, and each does its own work. The total of all of them, the **net work**, is the net change in the object's energy:

Key Equation: Work and energy

$$W_{\text{net}} = W_1 + W_2 + W_3 + \cdots = \Delta E. \tag{7.1}$$

If $W_{\text{net}} > 0$ the object ends up with more energy than it started with; if $W_{\text{net}} < 0$ it ends up with less; and if $W_{\text{net}} = 0$ the forces' contributions cancel and its energy is unchanged. Chapter 8 will show exactly which kind of energy changes (for a particle, it turns out to be the energy of its motion) and will give that energy a formula. For now, equation (7.1) is our definition: work is a change in energy.

Calculating the work

Only the part of a force *along* the direction of motion transfers energy. A force perpendicular to the motion, like the normal force on a block sliding across a level table, does no work at all, however large it is. For a constant force $\mathbf{F}$ acting while the object undergoes a straight-line displacement $\mathbf{d}$, the work is the component of the force parallel to the displacement, $F_{\parallel}$, times the magnitude of the displacement:

Key Equation: Work done by a constant force

$$W = F_{\parallel}\, d = F d \cos\theta = \mathbf{F} \cdot \mathbf{d}. \tag{7.2}$$

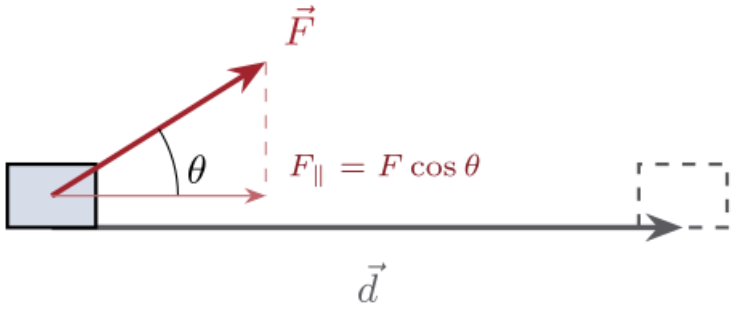


Figure 7.1. A constant force $\mathbf{F}$ acts on an object that undergoes a displacement $\mathbf{d}$ (the dashed outline shows the final position). Only the component of the force along the displacement, $F_{\parallel} = F\cos\theta$, does work; the work is $W = F_{\parallel} d$.

•Figure description

A constant force F applied at an angle theta to an object's horizontal displacement d, with a dashed outline showing the final position. Only the component of the force along the displacement does work.

The three forms say the same thing. In the first, $F_{\parallel}$ is the component of $\mathbf{F}$ along $\mathbf{d}$: positive if it points the same way as the displacement, negative if it points the opposite way. In the second, θ is the angle between $\mathbf{F}$ and $\mathbf{d}$ (Figure 7.1), so that $F_{\parallel} = F\cos\theta$. The third is the dot product of Chapter 3, which makes it clear that work is a scalar.

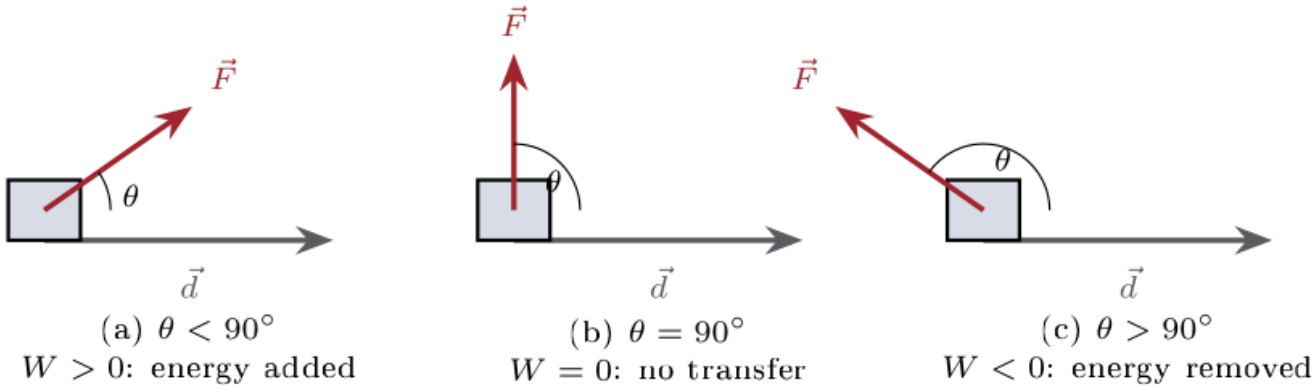


Figure 7.2. The sign of the work is carried entirely by $\cos\theta$. The displacement d is the same in all three cases; only the direction of the force differs. In (b) the force is perpendicular to the motion, like the normal force on a sliding block or gravity on a satellite in a circular orbit, and does no work at all.

•Figure description

Three panels showing the same displacement d with the force at a different angle in each: acute, so the work is positive; perpendicular, so the work is zero; and obtuse, so the work is negative. The sign of the work is carried entirely by cosine theta.

Note: Which angle?

The θ in $Fd\cos\theta$ is the angle between the force and the displacement, and that is often *not* the angle given in the problem. For a block sliding $2.0\ \mathrm{m}$ down a 30° incline, the angle given is the incline's, but the angle between the weight (straight down) and the displacement (down the slope) is 60°. The safer habit is to find $F_\parallel$ from the picture: the weight's component along the slope is $mg\sin 30^\circ$, pointing the way the block moves, so $W_g = (mg\sin 30^\circ)(2.0\ \mathrm{m})$. Both routes agree, because $\cos 60^\circ = \sin 30^\circ$, but only one of them depends on picking the right angle.

The SI unit of work is the **joule**: $1\ \mathrm{J} = 1\ \mathrm{N\cdot m} = 1\ \mathrm{kg\cdot m^2/s^2}$. A joule is about the work you do lifting an apple from the table to your mouth. The cosine carries the sign: a force with a component along the motion ($\theta < 90^\circ$) does positive work, a force with a component against the motion ($\theta > 90^\circ$) does negative work, and a force perpendicular to the motion ($\theta = 90^\circ$) does none (Figure 7.2).

When the force and displacement are given in components, the dot product is often the quicker route: $W = F_x d_x + F_y d_y + F_z d_z$. No angle is needed, and the sign comes out automatically.

Checkpoint 7.1

You lower a bucket into a well at constant speed. What is the sign of the work done on the bucket by the rope, and by gravity? What is the net work? Does the bucket's energy change?

Answer

Work and the free-body diagram

To find the work done by a specific force, you must know that force. These forces are given to you by Newton's laws, just like in Chapters 5 and 6. So a work problem is a Chapter 6 problem with one more step at the end: multiplying by a displacement.

Note: Strategy: finding the work done by each force

1. **Draw the free-body diagram** and mark the direction of the displacement d beside it.
2. **Find the unknown forces with Newton's laws.** Use the perpendicular equation for the normal force (and from it, friction $F_{f,k} = \mu_k F_N$); use the equation along the motion for a tension or an applied force. If the velocity is constant, $\mathbf{F}_{\mathrm{net}} = 0$.

3. **Find the angle between each force and** $\mathbf{d}$, and apply $W = Fd\cos\theta$. Forces perpendicular to $\mathbf{d}$ do no work; a force exactly opposite $\mathbf{d}$ does $-Fd$.
4. **Add the works** to get W_{net}, which is the change in the object's energy. As a check, W_{net} also equals the work done by the net force, $\mathbf{F}_{\text{net}} \cdot \mathbf{d}$.

Example 7.1: Work done by each force on a pushed block

A block of mass $2.50\ \text{kg}$ is pushed a distance $2.20\ \text{m}$ across a horizontal table by a constant force of $16.0\ \text{N}$ directed 25° below the horizontal. Friction between block and table is $4.0\ \text{N}$. Find the work done on the block by (a) the applied force, (b) friction, (c) the normal force, (d) gravity, and (e) the net force. (f) Does the block's energy increase or decrease, and by how much?

Solution

The displacement is horizontal, of magnitude $d = 2.20\ \text{m}$. For each force, the whole job is to find the angle θ between that force and the displacement, and then apply $W = Fd\cos\theta$. The cosine supplies the sign automatically.

(a) The applied force points 25° below the horizontal, and the displacement is horizontal, so the angle between them is 25°:

$$W_{\text{app}} = Fd\cos 25^\circ = (16.0\,\text{N})(2.20\,\text{m})(0.906) = 31.9\,\text{J}. \tag{7.1.1}$$

Only the horizontal component, $F\cos 25^\circ = 14.5\ \text{N}$, does work; the downward component, $F\sin 25^\circ$, is perpendicular to the motion and does none.

(b) Friction points backward, exactly opposite the displacement, so $\theta = 180^\circ$ and $\cos 180^\circ = -1$:

$$W_{\text{fric}} = F_f d\cos 180^\circ = (4.0\,\text{N})(2.20\,\text{m})(-1) = -8.8\,\text{J}. \tag{7.1.2}$$

The negative sign means friction takes energy away from the block.

(c, d) The normal force points up and gravity points down; both are perpendicular to the horizontal displacement, so $\theta = 90^\circ$ and $\cos 90^\circ = 0$. Each does *zero* work, however large it is:

$$W_N = 0, \qquad W_g = 0. \tag{7.1.3}$$

(e) Work is a scalar, so the net work is the ordinary sum of the four signed numbers:

$$W_{\text{net}} = W_{\text{app}} + W_{\text{fric}} + W_N + W_g = 31.9 + (-8.8) + 0 + 0 = \boxed{23.1\,\text{J}}. \tag{7.1.4}$$

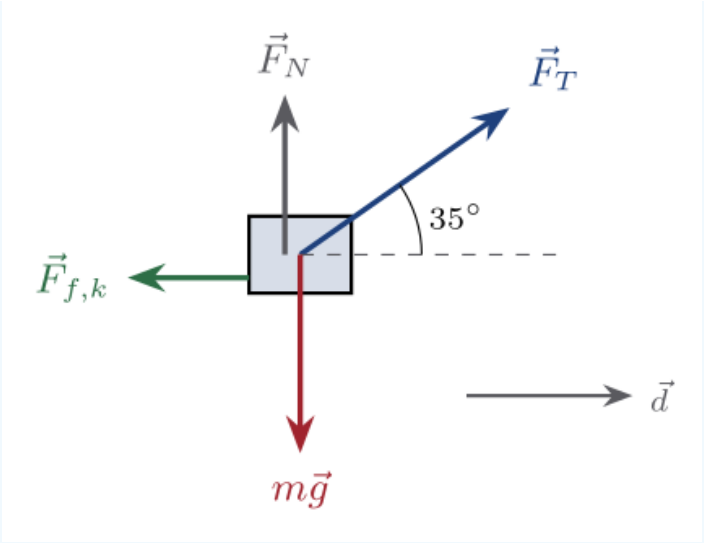
(f) By equation (7.1), the block's energy *increases* by $23.1\ \text{J}$. The push added $31.9\ \text{J}$; friction took $8.8\ \text{J}$ of it back out.

Check. Alternatively, find the net force first and then its work. The net horizontal force is $F\cos 25^\circ - F_f = 14.5 - 4.0 = 10.5\ \text{N}$ (vertically the forces balance, since the block stays on the table), and $(10.5\ \text{N})(2.20\ \text{m}) = 23.1\ \text{J}$. ✓ "Sum of the works" and "work of the sum" always agree.

Notice that the block's mass never entered the calculation, because friction was given to us directly. Had we been given the coefficient of friction instead, the mass would matter through the normal force. And because the push is directed *downward* at 25°, it presses the block into the table, so $F_N = mg + F\sin 25^\circ = 24.5 + 6.8 = 31.3\ \text{N}$, larger than the weight alone. The next example has to do exactly that.

Example 7.2: Pulling a sled at constant velocity

A child pulls a $20\ \text{kg}$ sled at constant velocity across level snow with a rope inclined 35° above the horizontal. The coefficient of kinetic friction between sled and snow is $\mu_k = 0.15$. The sled moves $12\ \text{m}$. (a) Find the tension in the rope. (b) Find the work done on the sled by each force acting on it. (c) What is the net work, and what happens to the sled's energy?



•Figure description
Free-body diagram of a sled pulled across level snow. The rope tension F sub T points up and to the right at 35 degrees above the horizontal; the normal force F sub N points up; the weight m g points down; kinetic friction points to the left, opposite the displacement d, which points to the right.

Solution

Step 1: Free-body diagram. Four forces act: the tension $\mathbf{F}_T$ at 35° above the horizontal, the normal force $\mathbf{F}_N$ up, the weight $m\mathrm{g}$ down, and kinetic friction $\mathbf{F}_{f,k}$ backward. Take x along the motion and y up. The velocity is constant, so $\sum F_x = 0$ and $\sum F_y = 0$.

Step 2: The normal force, from the vertical equation. The rope's upward component helps hold the sled up, so the snow need not push as hard:

$$F_N + F_T \sin 35^\circ - mg = 0 \quad \Longrightarrow \quad F_N = mg - F_T \sin 35^\circ. \tag{7.1.5}$$

F_T is still unknown, so we carry F_N as an expression.

Step 3: The tension, from the horizontal equation. Forward pull balances friction, and friction is $\mu_k F_N$:

$$F_T \cos 35^\circ = \mu_k \left(mg - F_T \sin 35^\circ\right). \tag{7.1.6}$$

Distribute μ_k, move the F_T term to the left, and factor out F_T:

$$F_T \cos 35^\circ + \mu_k F_T \sin 35^\circ = \mu_k mg \implies F_T = \frac{\mu_k mg}{\cos 35^\circ + \mu_k \sin 35^\circ}. \tag{7.1.7}$$

With $mg = (20)(9.8) = 196\ \mathrm{N}$,

$$F_T = \frac{(0.15)(196\,\mathrm{N})}{0.819 + (0.15)(0.574)} = \frac{29.4\,\mathrm{N}}{0.905} \approx \boxed{32.5\,\mathrm{N}}. \tag{7.1.8}$$

Then $F_N = 196 - (32.5)(0.574) = 177.4\ \mathrm{N}$ and $F_{f,k} = \mu_k F_N = (0.15)(177.4) = 26.6\ \mathrm{N}$.

Step 4: The work done by each force. The angle between $\mathbf{F}_T$ and the horizontal displacement is 35°; friction is opposite the displacement; $\mathbf{F}_N$ and $m\mathrm{g}$ are perpendicular to it:

$$\begin{aligned} W_T &= F_T d \cos 35^\circ = (32.5\,\mathrm{N})(12\,\mathrm{m})(0.819) = \boxed{319\,\mathrm{J}}, \\ W_f &= F_{f,k} d \cos 180^\circ = -(26.6\,\mathrm{N})(12\,\mathrm{m}) = \boxed{-319\,\mathrm{J}}, \\ W_N &= W_g = \boxed{0}. \end{aligned} \tag{7.1.9}$$

Step 5: Net work. $W_{\text{net}} = 319 - 319 + 0 + 0 = \boxed{0}$, so the sled's energy does not change. That is no accident: the velocity is

constant, so the net force is zero, and the work of a zero net force is zero. Every joule the rope puts into the sled, friction takes back out; it leaves the sled and warms the snow and the runners.

Check. The horizontal component of the pull, $F_T \cos 35^\circ = 26.6\ \text{N}$, equals the friction force, as constant velocity requires. ✓

The strategy also runs in reverse. If you know how much an object's energy should change, equation (7.1) tells you how much net work is needed, and $W_{\text{net}} = F_{\text{net}}\ d$ (for a constant net force along the motion) tells you how far the object must move to get it.

Example 7.3: How far to drag a crate

A worker drags a $40\ \text{kg}$ crate across a level floor by pulling on a rope with a force of $150\ \text{N}$ directed 30° above the horizontal. The coefficient of kinetic friction is $\mu_k = 0.30$. Over what distance must the crate be dragged for its energy to increase by $500\ \text{J}$? How much work does the rope do over that distance?

Solution

Step 1: The normal force and friction. The free-body diagram is the sled's, with different numbers. Vertically there is no acceleration:

$$F_N = mg - F_T \sin 30^\circ = (40)(9.8) - (150)(0.500) = 392 - 75 = 317\,\text{N}, \tag{7.1.10}$$

so $F_{f,k} = \mu_k F_N = (0.30)(317) = 95.1\ \text{N}$.

Step 2: The net force. Horizontally, $F_{\text{net}} = F_T \cos 30^\circ - F_{f,k} = 129.9 - 95.1 = 34.8\ \text{N}$, forward. Unlike the sled, this crate is *not* in equilibrium: the forward pull exceeds the friction, and the net force does positive work.

Step 3: Solve for the distance. The net force is constant and along the displacement, so $W_{\text{net}} = F_{\text{net}}\ d$. Setting $W_{\text{net}} = \Delta E$ and dividing both sides by F_{net}:

$$d = \frac{\Delta E}{F_{\text{net}}} = \frac{500\,\text{J}}{34.8\,\text{N}} \approx \boxed{14.4\,\text{m}}. \tag{7.1.11}$$

(A joule divided by a newton is a meter.)

Step 4: Work done by the rope. Over this distance,

$$W_T = F_T d \cos 30^\circ = (150\,\text{N})(14.4\,\text{m})(0.866) \approx \boxed{1.87 \times 10^3\,\text{J}}. \tag{7.1.12}$$

Check. Friction does $W_f = -(95.1)(14.4) = -1.37 \times 10^3\ \text{J}$, and $1870 - 1370 = 500\ \text{J}$. ✓ Most of the rope's work is undone by friction; only about a quarter of it remains as an increase in the crate's energy.

Answers to Checkpoints

Checkpoint 7.1. The rope pulls up while the bucket moves down, so $\theta = 180^\circ$ and the rope does *negative* work. Gravity points down, along the motion, so it does *positive* work. At constant speed the rope's pull equals the bucket's weight, so the two works are equal in size and opposite in sign: the net work is zero, and by $W_{\text{net}} = \Delta E$ the bucket's energy does not change. (Constant velocity always means zero net force, and so zero net work.)

This page titled 7.1: Work Done by a Constant Force is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.2: Work Done by Gravity

Near the Earth's surface the gravitational force on an object is the constant force $m\mathbf{g}$, directed straight down. Because it is constant, equation (7.2) applies, and the result is so simple and so useful that it deserves its own section.

Take y upward. If the object's displacement is $\mathbf{d} = d_x\ \hat{\imath} + d_y\ \hat{\jmath}$ and the gravitational force is $m\mathbf{g} = -mg\ \hat{\jmath}$, the dot product picks out only the vertical part of the displacement:

$$W_g = (m\mathbf{g}) \cdot \mathbf{d} = (0)(d_x) + (-mg)(d_y) = -mg\, d_y. \tag{7.2.1}$$

The vertical component of the displacement is just the change in height, $d_y = \Delta y = y_f - y_i$. So:

Key Equation: Work done by gravity near Earth's surface

$$W_g = -mg\,\Delta y = -mg\,(y_f - y_i)\,. \tag{7.3}$$

An object that moves *down* ($\Delta y < 0$) has positive work done on it by gravity; one that moves *up* has negative work done on it. Horizontal motion contributes nothing.

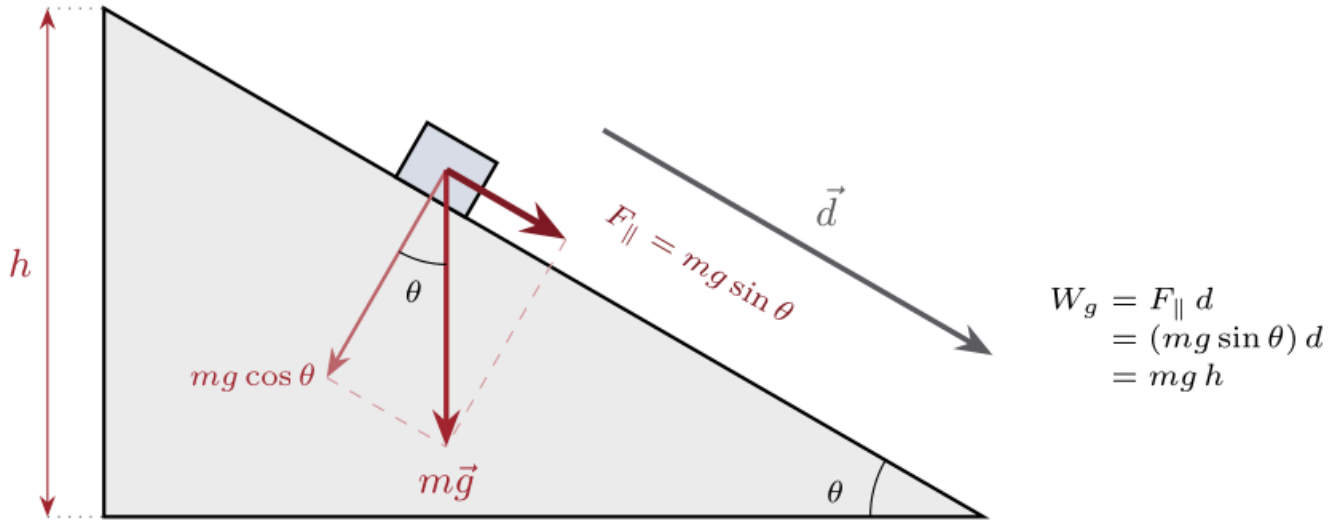


Figure 7.3. Gravity's work on a block sliding down an incline. The weight $m\mathbf{g}$ splits into a component $mg\cos\theta$ perpendicular to the slope, which does no work, and a component $F_\parallel = mg\sin\theta$ parallel to the displacement $\mathbf{d}$. As in Figure 7.1, only $F_\parallel$ does work: $W_g = F_\parallel d = (mg\sin\theta)\ d$. Since $d\sin\theta = h$, the work is mgh, the same as for a straight drop through the same height.

•Figure description

A block sliding a distance d down an incline of angle theta, dropping a vertical height h. The block's weight m g points straight down and is resolved into two components: m g cosine theta perpendicular to the slope, and F sub parallel equals m g sine theta down the slope, parallel to the displacement d. The angle theta appears between m g and the perpendicular component. Only F sub parallel does work, so W sub g equals F sub parallel times d, which equals m g sine theta times d, which equals m g h.

The striking thing about this result is what it leaves out. The horizontal part of the motion does not appear, and neither does the path. Slide a box down a long gentle ramp or drop it straight down through the same height h: gravity does $+mgh$ either way. Figure 7.3 shows why on an incline. Along the ramp, the component of gravity in the direction of motion is $mg\sin\theta$, which is smaller than mg, but it acts over the longer distance $d = h/\sin\theta$, and the product is the same:

$$W_g = (mg\sin\theta)\,d = (mg\sin\theta)\,\frac{h}{\sin\theta} = mgh. \tag{7.2.2}$$

Any path, however curved, can be built from many short inclines, so the same is true of every path. Gravity is not the only force with this property, and in Chapter 8 we will see that the forces that share it form a special class: the ones for which a "stored"

energy can be defined.

Lifting and lowering

A common special case is lifting. To raise an object a height h at constant speed, your hand must push up with a force equal to the weight, mg (the net force is zero), in the same direction as the displacement. Your hand therefore does work

$$W_{\text{hand}} = +mgh, \tag{7.4}$$

while gravity, opposing the upward displacement, does $W_g = -mgh$. The net work is zero, and the object's energy, by equation (7.1), is unchanged. Lowering the object at constant speed reverses both signs.

This is worth a moment's thought, because it seems to say that your effort has vanished: you did mgh of work, and the object has no more energy to show for it. In fact something *has* changed: the object is now high up, where it can fall and gather speed that it could not have gathered on the floor. Keeping track of that requires thinking about the object and the Earth together, and that is the first job of Chapter 8. For now, the bookkeeping of equation (7.1) for the object alone is exactly right: zero net work, no change in its energy.

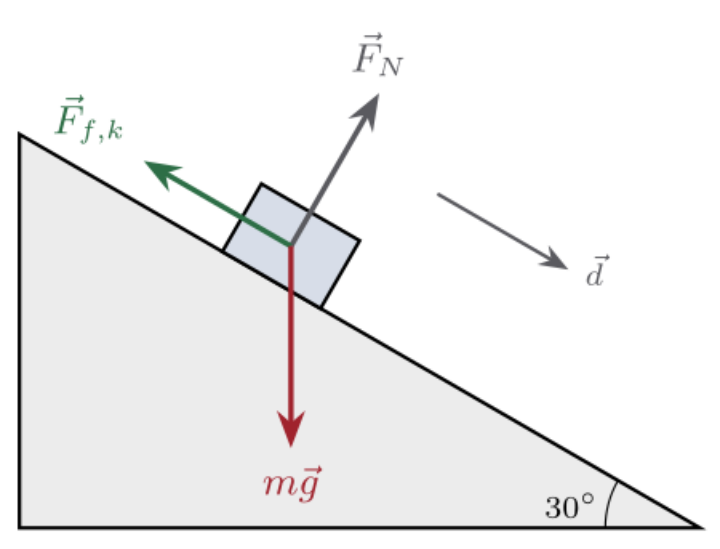
Checkpoint 7.2

You pick up a heavy box, carry it across the room at constant speed, and set it down on a second table exactly as high as the first. Over the whole trip, how much work did gravity do on the box? How much net work was done on it? Why, then, are you tired?

Answer

Example 7.4: A block sliding down a rough ramp

A $5.0\ \text{kg}$ block slides $3.0\ \text{m}$ down a ramp inclined at 30°. The coefficient of kinetic friction between the block and the ramp is $\mu_k = 0.25$. Find the work done on the block by (a) gravity, (b) the normal force, and (c) friction. (d) What is the net work, and how does the block's energy change?



•Figure description
Free-body diagram of a block sliding down an incline. The normal force F sub N points perpendicular to the surface, the weight m g points straight down, and kinetic friction points up the slope, opposite the displacement d, which points down the slope.

Solution

Step 1: Forces from Newton's laws. Tilt the axes as in Chapter 6: x down the slope (the direction of d), y perpendicular to it. The block does not accelerate perpendicular to the slope, so

$$F_N = mg\cos 30^\circ = (5.0)(9.8)(0.866) = 42.4\,\mathrm{N}, \tag{7.2.3}$$

so the friction force is $F_{f,k} = \mu_k F_N = (0.25)(42.4) = 10.6\ \mathrm{N}$.

Step 2: Work done by gravity. The block drops a height $h = d\sin 30^\circ = (3.0)(0.500) = 1.5\ \mathrm{m}$, so $\Delta y = -1.5\ \mathrm{m}$ and, by equation (7.3),

$$W_g = -mg\,\Delta y = -(5.0)(9.8)(-1.5) = \boxed{+73.5\,\mathrm{J}}. \tag{7.2.4}$$

(Using the component along the slope instead: $(mg\sin 30^\circ)\ d = (24.5)(3.0) = 73.5\ \mathrm{J}$, the same.)

Step 3: The normal force and friction. The normal force is perpendicular to the slope, and so to the displacement: $W_N = \boxed{0}$. Friction points up the slope, opposite the displacement:

$$W_f = -F_{f,k}d = -(10.6\,\mathrm{N})(3.0\,\mathrm{m}) = \boxed{-31.8\,\mathrm{J}}. \tag{7.2.5}$$

Step 4: Net work. $W_{\text{net}} = 73.5 + 0 - 31.8 = \boxed{41.7\,\mathrm{J}}$. The block's energy increases by $41.7\ \mathrm{J}$: gravity puts energy in faster than friction takes it out.

Check. The net force along the slope is $mg\sin 30^\circ - F_{f,k} = 24.5 - 10.6 = 13.9\ \mathrm{N}$, and $(13.9)(3.0) = 41.7\ \mathrm{J}$. ✓

Example 7.5: Hoisting a beam

A crane lifts a $500\ \mathrm{kg}$ steel beam a height of $12\ \mathrm{m}$, with a constant upward acceleration of $0.50\ \mathrm{m/s^2}$. Find (a) the tension in the cable, (b) the work done by the cable, (c) the work done by gravity, and (d) the change in the beam's energy.

Solution

(a) Two forces act on the beam: the tension F_T up and the weight mg down. Newton's second law, with up positive, gives $F_T - mg = ma$, so

$$F_T = m(g + a) = (500\,\mathrm{kg})(9.8 + 0.50)\,\mathrm{m/s^2} = \boxed{5150\,\mathrm{N}}. \tag{7.2.6}$$

(b) The tension is along the upward displacement, and (c) the beam rises $\Delta y = +12\ \mathrm{m}$, so

$$\begin{aligned} W_T &= F_T d = (5150\,\mathrm{N})(12\,\mathrm{m}) = \boxed{6.18\times 10^4\,\mathrm{J}}, \\ W_g &= -mg\,\Delta y = -(500)(9.8)(12)\,\mathrm{J} = \boxed{-5.88\times 10^4\,\mathrm{J}}. \end{aligned} \tag{7.2.7}$$

(d) The beam's energy changes by the net work:

$$\Delta E = W_{\text{net}} = 61\,800 - 58\,800 = \boxed{3.0\times 10^3\,\mathrm{J}}. \tag{7.2.8}$$

Check. The net force is $ma = (500)(0.50) = 250\ \mathrm{N}$, and $(250)(12) = 3000\ \mathrm{J}$. ✓ Notice how lopsided the bookkeeping is. Almost all of the cable's work ($58.8\ \mathrm{kJ}$ of $61.8\ \mathrm{kJ}$) goes into fighting gravity; only $3.0\ \mathrm{kJ}$ is left over as a change in the beam's energy. Had the beam been lifted at constant speed, the cable would have done exactly $58.8\ \mathrm{kJ}$ and the beam's energy would not have changed at all.

Example 7.6: A skier on a gentle slope

A $60\ \mathrm{kg}$ skier starts down a 15° slope. The coefficient of kinetic friction between skis and snow is 0.080. How far down the slope must she travel for her energy to increase by $2.0\ \mathrm{kJ}$?

Solution

Step 1: Net force along the slope. With axes along and perpendicular to the slope, $F_N = mg\cos 15^\circ$, so friction is $\mu_k mg\cos 15^\circ$, up the slope. Gravity's component down the slope is $mg\sin 15^\circ$. The net force along the slope is therefore

$$F_{\text{net}} = mg\sin 15^\circ - \mu_k mg\cos 15^\circ = mg\left(\sin 15^\circ - \mu_k \cos 15^\circ\right). \tag{7.2.9}$$

With $mg = 588\ \text{N}$, this is $(588)(0.2588 - 0.0773) = (588)(0.1815) = 106.7\ \text{N}$.

Step 2: Solve for the distance. $W_{\text{net}} = F_{\text{net}}\ d = \Delta E$, so

$$d = \frac{\Delta E}{F_{\text{net}}} = \frac{2000\,\text{J}}{106.7\,\text{N}} \approx \boxed{19\,\text{m}}. \tag{7.2.10}$$

Check. Over $18.7\ \text{m}$ of slope she drops $h = (18.7)\sin 15^\circ = 4.85\ \text{m}$, so gravity does $mgh = (588)(4.85) = 2.85\ \text{kJ}$, and friction does $-(45.4\ \text{N})(18.7\ \text{m}) = -0.85\ \text{kJ}$. The sum is $2.0\ \text{kJ}$. ✓

Answers to Checkpoints

Checkpoint 7.2. Gravity does zero work over the whole trip: $-mgh$ while you lift the box and $+mgh$ while you lower it, and nothing on the level carry, since gravity is perpendicular to horizontal motion. Equivalently, $\Delta y = 0$ overall, so $W_g = -mg\ \Delta y = 0$. The box starts and ends at rest, and at constant speed in between, so the net work on it is zero too. You are tired because your muscles use chemical energy even when they do no net work on the box: holding a load steady requires muscle fibers to contract and relax continually, and that energy ends up as heat in your body. "Work" in physics is a narrower idea than effort.

This page titled 7.2: Work Done by Gravity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.3: Work Done by a General Force

Equation (7.2) assumes the force is constant. Many forces are not. The most important example is a spring: the more you stretch it, the harder it pulls back. Within its elastic range, the force a spring exerts is described by **Hooke's law**,

Key Equation: Hooke's law

$$F_s = -kx, \tag{7.5}$$

where x is the displacement of the spring's end from its relaxed position and k is the **spring constant** (in $\mathrm{N/m}$). A stiff spring has a large k. The minus sign indicates that the spring force always opposes the stretch or compression: pull the end in $+x$ and the spring pulls back in $-x$; push it in $-x$ and the spring pushes back in $+x$. Because this force changes as the object moves, we cannot use $W = Fd$.

The work integral

When the force varies with position, we divide the displacement into pieces so small that the force is nearly constant over each, compute the work $F\ dx$ done on each piece, and add them up. That is, we integrate:

Key Equation: Work done by a variable force

$$W = \int_{x_i}^{x_f} F(x)\,dx. \tag{7.6}$$

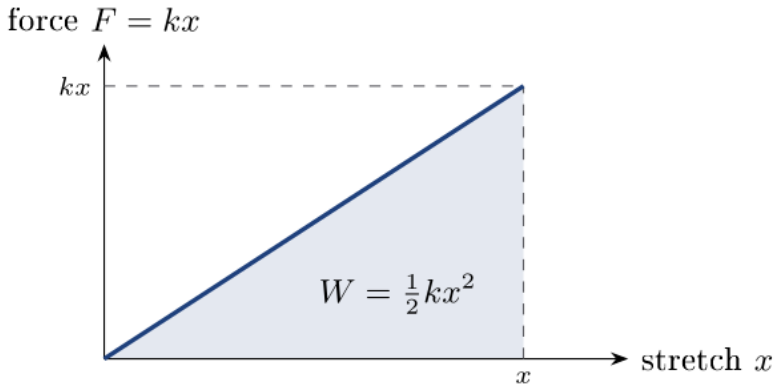


Figure 7.4. For a spring, the force needed to stretch it rises linearly with stretch, $F = kx$. The work done is the shaded triangular area under the line: base x, height kx, so area $\frac{1}{2}(x)(kx) = \frac{1}{2}kx^2$.

•Figure description

A graph of spring force versus stretch, a straight line through the origin of slope k. The area under the line from zero to x is shaded as a triangle of base x and height k x, giving work equal to one half k x squared.

Geometrically, the work is the area under the force-versus-position graph (Figure 7.4), with area below the axis counted as negative. For a constant force the "area" is a rectangle, $F\ (x_f - x_i)$, and equation (7.6) reduces to $W = Fd$.

The same idea works for motion along any path, curved or straight, in two or three dimensions. Over each tiny displacement $d\mathbf{r}$ the force does work $\mathbf{F} \cdot d\mathbf{r}$, and the total is the sum of all of them, $W = \int \mathbf{F} \cdot d\mathbf{r}$. In this book we will need the general form only occasionally; equation (7.6), for motion along a line, is the one to know.

Example 7.7: Work done by expanding gas on a bullet

A $100\ \text{g}$ bullet is fired down a rifle barrel $0.600\ \text{m}$ long. With the origin at the bullet's starting point, the expanding gas exerts a force (in newtons, x in meters)

$$F(x) = 15000 + 10000x - 25000x^2. \tag{7.3.1}$$

Find the work the gas does on the bullet over the length of the barrel. By how much does it change the bullet's energy?

Solution

Step 1: Set up the integral. The force varies with position, so $W = Fd$ does not apply; we integrate the force from $x = 0$ to $x = 0.600\ \text{m}$ (equation (7.6)):

$$W = \int_0^{0.600} \left(15000 + 10000x - 25000x^2\right) dx. \tag{7.3.2}$$

Step 2: Integrate term by term. The integral of a sum is the sum of the integrals, and each term follows the power rule, $\int x^n\ dx = x^{n+1}/(n+1)$: raise the power by one and divide by the new power. The constant 15000 is $15000x^0$, so it becomes $15000x$; the $10000x$ term becomes $10000\ x^2/2 = 5000x^2$; and the $-25000x^2$ term becomes $-25000\ x^3/3$:

$$W = \left[15000x + 5000x^2 - \tfrac{25000}{3}x^3\right]_0^{0.600}. \tag{7.3.3}$$

Step 3: Evaluate at the limits. Substitute the upper limit and subtract the value at the lower limit. At $x = 0$ every term is zero, so only the upper limit contributes. Work out the powers first: $(0.600)^2 = 0.360$ and $(0.600)^3 = 0.216$. Then

$$W = 15000(0.600) + 5000(0.360) - \tfrac{25000}{3}(0.216) = 9000 + 1800 - 1800 = \boxed{9.00 \times 10^3\ \text{J}}. \tag{7.3.4}$$

(For the last term, $25000 \times 0.216 = 5400$, and $5400/3 = 1800$.) The second and third terms happen to cancel exactly here: the force's early rise and later fall contribute equal and opposite amounts relative to the constant $15000\ \text{N}$, so the gas does the same work a steady $15000\ \text{N}$ push would have done over $0.600\ \text{m}$.

Energy change. Treating the gas as the only force that does appreciable work on the bullet (gravity is perpendicular to a level barrel, and friction in the barrel is small by comparison), the bullet's energy increases by $9.00\ \text{kJ}$. Notice that the bullet's mass was not needed. It will matter in Chapter 8, when we find out how fast an object carrying this much energy of motion must be moving.

Work done by a spring

Applying equation (7.6) to a spring calls for care with signs, and this is worth slowing down for. There are always *two* forces to think about: the force the *spring* exerts on whatever is attached to it, $F_s = -kx$, and the force *you* (or whatever is stretching it) exert on the spring, which, if the stretching is done slowly, is equal and opposite, $F = +kx$.

Work done by you. As you stretch the spring from its relaxed length to a stretch x, your force $+kx$ points along the motion, and you do positive work:

$$W_{\text{you}} = \int_0^x kx'\ dx' = \tfrac{1}{2}kx^2. \tag{7.7}$$

(The primed variable x' is just a placeholder that runs over the interval, so that the upper limit x is not confused with the variable of integration.) Compressing the spring a distance x takes the same work, since only x^2 appears.

Work done by the spring. The spring's own force is the negative of yours, so over the same stretch it does $-\frac{1}{2}kx^2$. More generally, as the end of the spring moves from x_i to x_f,

$$W_s = \int_{x_i}^{x_f} (-kx)\, dx = -\left[\tfrac{1}{2}kx^2\right]_{x_i}^{x_f}, \tag{7.3.5}$$

which gives

Key Equation: Work done by a spring

$$W_s = \tfrac{1}{2}kx_i^2 - \tfrac{1}{2}kx_f^2. \tag{7.8}$$

Read the signs carefully. If the spring ends up more stretched or compressed than it started ($|x_f| > |x_i|$), it does *negative* work on the object attached to it: it resists being deformed. If it ends up closer to its relaxed length, it does *positive* work: it pushes or pulls the object back toward $x = 0$, adding energy as it goes. And, like gravity, the spring's work depends only on where the end starts and finishes, not on how it got there.

The work grows as the *square* of the stretch: stretching a spring twice as far takes four times the work. In Chapter 8 we will see that this work, like the work done lifting an object, is not lost; it is stored in the spring.

Checkpoint 7.3

A spring is compressed $0.10\ \text{m}$. In which direction does the spring force point, and what is the sign of the work the spring does on the hand compressing it? Now the hand releases the spring and it pushes back to its relaxed length: what is the sign of the work the spring does on the hand over that second stage?

Answer

Example 7.8: Drawing a bow

An archer pulls a bowstring back $0.400\ \text{m}$ by exerting a force that increases uniformly from zero to $230\ \text{N}$. (a) What is the equivalent spring constant of the bow? (b) How much work does the archer do in drawing the bow?

Solution

(a) Because the force rises uniformly with displacement, the bow behaves like a spring, $F = kx$. At full draw, $F = 230\ \text{N}$ when $x = 0.400\ \text{m}$. Divide both sides of $F = kx$ by x:

$$k = \frac{F}{x} = \frac{230\,\text{N}}{0.400\,\text{m}} = 575\,\text{N/m}. \tag{7.3.6}$$

(b) The work is $\frac{1}{2}kx^2$ (equation (7.7)). Square the displacement first, then multiply:

$$W = \tfrac{1}{2}kx^2 = \tfrac{1}{2}(575\,\text{N/m})(0.400\,\text{m})^2 = \tfrac{1}{2}(575\,\text{N/m})(0.160\,\text{m}^2) = \boxed{46.0\,\text{J}}. \tag{7.3.7}$$

The units combine as $(\ \text{N/m})(\ \text{m}^2) =\ \text{N}\cdot\text{m} =\ \text{J}$.

Check. The work is also the area under the force–displacement line in Figure 7.4: a triangle with base $0.400\ \text{m}$ and height $230\ \text{N}$, so $W = \frac{1}{2}(0.400)(230) = 46.0\ \text{J}$. This route does not even need k. It also shows why the answer is *half* of $(230\ \text{N})(0.400\ \text{m})$: the archer's force is $230\ \text{N}$ only at full draw, and averages half that over the pull.

Example 7.9: Lowering a mass onto a spring

A $1.5\ \text{kg}$ mass is hooked onto the end of a vertical spring ($k = 120\ \text{N/m}$) hanging from the ceiling. Starting from the spring's relaxed length, you support the mass with your hand and lower it slowly until it hangs at rest on its own. (a) How far does the spring stretch? Find the work done on the mass by (b) gravity, (c) the spring, and (d) your hand. (e) What is the net work?

Solution

(a) When the mass hangs at rest by itself, only two forces act on it: the spring pulling up with kd and gravity pulling down with mg. Equilibrium requires $kd = mg$, so

$$d = \frac{mg}{k} = \frac{(1.5)(9.8)\,\mathrm{N}}{120\,\mathrm{N/m}} = \frac{14.7\,\mathrm{N}}{120\,\mathrm{N/m}} = \boxed{0.1225\,\mathrm{m}}. \tag{7.3.8}$$

(b) The mass moves down d, so $\Delta y = -d$ and $W_g = -mg\ \Delta y = mgd = (14.7)(0.1225) = \boxed{1.80\,\mathrm{J}}$.

(c) The spring goes from $x_i = 0$ to $x_f = d$, so by equation (7.8)

$$W_s = 0 - \tfrac{1}{2}kd^2 = -\tfrac{1}{2}(120)(0.1225)^2 = \boxed{-0.900\,\mathrm{J}}. \tag{7.3.9}$$

(d) Lowering slowly means the mass is essentially in equilibrium the whole way down, so at each position your hand supplies whatever upward force is needed to balance the rest: $F_{\text{hand}} = mg - kx$. This force varies, so integrate. It points up while the mass moves down, so its work is negative:

$$W_{\text{hand}} = -\int_0^d (mg - kx)\,dx = -\left(mgd - \tfrac{1}{2}kd^2\right) = -(1.80 - 0.900) = \boxed{-0.900\,\mathrm{J}}. \tag{7.3.10}$$

(e) $W_{\text{net}} = 1.80 - 0.900 - 0.900 = \boxed{0}$, as it must be: the mass starts and ends at rest, held in equilibrium throughout, so its energy does not change.

The result splits evenly in an interesting way. Of the $1.80\ \mathrm{J}$ that gravity supplied, the spring took half and your hand took the other half. Had you let go at the top instead of lowering the mass gently, your hand would have done no work at all, and the $0.900\ \mathrm{J}$ it would otherwise have absorbed would have nowhere to go but into motion. The mass would overshoot and bounce, a situation Chapter 8 is designed to handle.

Answers to Checkpoints

Checkpoint 7.3. A compressed spring pushes outward, toward its relaxed length, that is, against the hand. While the hand compresses it, the hand moves inward against that outward force, so the spring does *negative* work on the hand (the hand does positive work on the spring). In equation (7.8), $x_i = 0$ and $|x_f| = 0.10\ \mathrm{m}$, so $W_s < 0$. During the release the hand moves outward in the same direction as the spring's push, so the spring does *positive* work on the hand; now $|x_i| = 0.10\ \mathrm{m}$ and $x_f = 0$, and $W_s > 0$.

This page titled 7.3: Work Done by a General Force is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.4: Power

Two motors that each raise the same load the same height do the same work, but the one that does it in half the time is, in everyday language, twice as powerful. **Power** measures how fast work is done, or equivalently how fast energy is transferred. The **average power** over a time interval Δt is the work done divided by the time, and the **instantaneous power** is its limit as $\Delta t \to 0$:

Key Equation: Power

$$P_{\text{avg}} = \frac{W}{\Delta t}, \qquad P = \frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v}. \tag{7.9}$$

The SI unit of power is the **watt**: $1\ \text{W} = 1\ \text{J/s}$. (The familiar horsepower is just another unit of power: $1\ \text{hp} \approx 746\ \text{W}$.) The compact form $P = \mathbf{F} \cdot \mathbf{v}$ follows from the definition of work: in a short time dt the object moves $d\mathbf{r} = \mathbf{v}\ dt$, so a force $\mathbf{F}$ does work $dW = \mathbf{F} \cdot \mathbf{v}\ dt$, and dividing by dt gives the power. For a force along the velocity it is simply $P = Fv$. It is often the easiest way to find the power a force delivers to an object moving at a known velocity.

Power brings *time* into problems that otherwise have none. Rearranged, equation (7.9) answers "how long?": a motor that can deliver power P needs at least $\Delta t = W/P$ to do work W. Rearranged the other way, $v = P/F$, it answers "how fast?": a machine delivering power P against a force F can move at most at speed P/F. Electric companies bill you for energy in a unit built on exactly this idea: one *kilowatt-hour* is the work done in one hour at a power of one kilowatt, $(1000\ \text{J/s})(3600\ \text{s}) = 3.6 \times 10^6\ \text{J}$.

Checkpoint 7.4

A car cruises at a steady $30\ \text{m/s}$ on a level highway against $500\ \text{N}$ of total resistance. What power does the engine deliver? If the driver doubles the speed and the resistive force quadruples (as drag does at highway speeds), by what factor does the required power change?

Answer

Example 7.10: Power to run an elevator

An elevator car of mass $1600\ \text{kg}$ carries passengers of total mass $200\ \text{kg}$, and a constant friction force of $4000\ \text{N}$ opposes its motion. (a) What power must the motor deliver to lift the car at a constant $3.00\ \text{m/s}$? (b) What power must it deliver at the instant the speed is v if the car is to accelerate upward at $1.00\ \text{m/s}^2$?

Solution

The total mass is $m = 1600 + 200 = 1800\ \text{kg}$, so the total weight is $mg = (1800\ \text{kg})(9.8\ \text{m/s}^2) = 17640\ \text{N}$.

The plan. The motor delivers its power through the cable, so we need the cable tension F_T and then $P = \mathbf{F}_T \cdot \mathbf{v}$ from equation (7.9). The tension and the velocity both point up, so the angle between them is zero, $\cos 0 = 1$, and the dot product is simply $P = F_T v$.

(a) Three forces act on the car: the tension F_T up, the weight mg down, and friction F_f down (opposing the upward motion). At constant velocity the acceleration is zero, so taking up as positive,

$$F_T - mg - F_f = 0 \quad \Longrightarrow \quad F_T = mg + F_f = 17640 + 4000 = 21640\,\text{N}. \tag{7.4.1}$$

The cable moves with the car at $3.00\ \text{m/s}$, so

$$P = F_T v = (21640\,\text{N})(3.00\,\text{m/s}) \approx \boxed{6.49 \times 10^4\,\text{W}} \approx 65\,\text{kW}. \tag{7.4.2}$$

(The units: $\text{N} \cdot \text{m/s} = \text{J/s} = \text{W}$.)

(b) Now the acceleration is $a = +1.00\ \text{m/s}^2$, so the right side of Newton's second law is ma instead of zero. Add mg and F_f to both sides, then factor m out of the two terms that contain it:

$$F_T - mg - F_f = ma \implies F_T = mg + ma + F_f = m(g + a) + F_f. \tag{7.4.3}$$

Substituting, $F_T = (1800)(9.8 + 1.00) + 4000 = (1800)(10.8) + 4000 = 19440 + 4000 = 23440\ \text{N}$. The extra $1800\ \text{N}$ beyond part (a) is exactly ma, the force needed to accelerate the car. The power then depends on the instantaneous speed:

$$P = F_T v = \boxed{(2.34 \times 10^4\,\text{N})\, v}. \tag{7.4.4}$$

Unlike part (a), the answer is a formula rather than a single number. That is not an oversight: while the car accelerates its speed is changing every instant, and since $P = F_T v$ with F_T fixed, the power the motor must deliver climbs steadily as the car speeds up. At $v = 3.00\ \text{m/s}$, for instance, it is $(23440)(3.00) \approx 70\ \text{kW}$; by $6.00\ \text{m/s}$ it has doubled to $141\ \text{kW}$.

Example 7.11: A winch on a loading ramp

A winch drags an $80\ \text{kg}$ crate up a ramp inclined at 20°, pulling on a cable parallel to the ramp. The coefficient of kinetic friction is 0.30, and the crate moves at a constant $0.50\ \text{m/s}$. (a) What power does the winch deliver? (b) How long does it take to haul the crate up the $6.0\ \text{m}$ ramp, and how much work does the winch do in that time? (c) If the winch can deliver at most $1.0\ \text{kW}$, what is the fastest it can haul the crate?

Solution

Step 1: The cable tension. Use axes along and perpendicular to the ramp. Perpendicular to the ramp, $F_N = mg\cos 20^\circ$, so friction is $\mu_k mg\cos 20^\circ$, directed *down* the ramp (the crate moves up). Along the ramp the velocity is constant, so the tension balances gravity's down-slope component plus friction:

$$F_T = mg\sin 20^\circ + \mu_k mg\cos 20^\circ = mg\left(\sin 20^\circ + \mu_k \cos 20^\circ\right). \tag{7.4.5}$$

With $mg = 784\ \text{N}$: $F_T = 784\ (0.342 + 0.282) = (784)(0.624) = 489\ \text{N}$.

Step 2: (a) Power. The cable pulls along the velocity, so

$$P = F_T v = (489\,\text{N})(0.50\,\text{m/s}) \approx \boxed{245\,\text{W}}. \tag{7.4.6}$$

Step 3: (b) Time and work. At constant speed, $\Delta t = d/v = (6.0\ \text{m})/(0.50\ \text{m/s}) = \boxed{12\,\text{s}}$. The work is the power times the time:

$$W = P\,\Delta t = (245\,\text{W})(12\,\text{s}) \approx \boxed{2.9 \times 10^3\,\text{J}}. \tag{7.4.7}$$

(Directly, $W = F_T d = (489)(6.0) = 2.9 \times 10^3\ \text{J}$, the same.)

Step 4: (c) Top speed. The tension needed does not depend on the speed (kinetic friction does not), so with $P = F_T v$ the speed is at most

$$v_{\max} = \frac{P_{\max}}{F_T} = \frac{1000\,\text{W}}{489\,\text{N}} \approx \boxed{2.0\,\text{m/s}}, \tag{7.4.8}$$

and the trip up the ramp would then take $6.0/2.04 \approx 2.9\ \text{s}$. A more powerful winch does not do any more work (the $2.9\ \text{kJ}$ is fixed by the forces and the ramp); it just does it faster.

Example 7.12: Climbing a hill at highway speed

A $1400\ \text{kg}$ car climbs a long, straight 5.0° hill at a constant $25\ \text{m/s}$ against a total resistive force (air drag plus rolling friction) of $600\ \text{N}$. (a) What power must the engine deliver to the wheels? (b) How long does it take the car to gain $50\ \text{m}$ of elevation, and how much of the engine's work over that time goes into fighting gravity rather than resistance?

Solution

(a) Along the road the velocity is constant, so the driving force from the road on the tires balances the down-slope component of gravity plus the resistance:

$$F_{\text{drive}} = mg\sin 5.0^\circ + R = (1400)(9.8)(0.0872) + 600 = 1196 + 600 = 1796\,\text{N}. \tag{7.4.9}$$

This force is along the velocity, so

$$P = F_{\text{drive}}\,v = (1796\,\text{N})(25\,\text{m/s}) \approx \boxed{4.5\times10^4\,\text{W}} = 45\,\text{kW} \approx 60\,\text{hp}. \tag{7.4.10}$$

(b) To gain $50\ \text{m}$ of height the car must travel $L = 50/\sin 5.0^\circ = 574\ \text{m}$ along the road, which at $25\ \text{m/s}$ takes

$$\Delta t = \frac{L}{v} = \frac{574\,\text{m}}{25\,\text{m/s}} \approx \boxed{23\,\text{s}}. \tag{7.4.11}$$

Over that distance gravity does $W_g = -mg\ \Delta y = -(1400)(9.8)(50) = -6.86\times10^5$ J, and resistance does $-(600)(574) = -3.44\times10^5$ J. The engine's work, $(1796)(574) = 1.03\times10^6$ J, cancels both (the car's speed, and so its energy, is unchanged). The fraction spent on gravity is $686/1030 \approx \boxed{67\%}$. On the level the same car at the same speed would need only $(600)(25) = 15$ kW: two-thirds of the power on this hill goes into climbing.

Answers to Checkpoints

Checkpoint 7.4. At constant speed the engine's forward force equals the 500 N resistance, so $P = Fv = (500\ \text{N})(30\ \text{m/s}) = 1.5\times10^4$ W $= 15$ kW. With $P = Fv$, doubling v and quadrupling F multiplies the power by $2\times4 = 8$, to 120 kW. This steep rise, with power growing as the *cube* of the speed, is why highway fuel economy falls off so quickly at high speed.

This page titled 7.4: Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.S: Work and Power (Summary)

Summary

- **Work** is the energy a force transfers to an object (positive work) or from it (negative work). The **net work**, the sum of the works done by all the forces, is the change in the object's energy, $W_{\text{net}} = \Delta E$. What forms that energy takes is the subject of Chapter 8.
- For a constant force, $W = Fd\cos\theta = \mathbf{F}\cdot\mathbf{d}$. Only the component of force along the displacement does work; a perpendicular force (such as a normal force on level ground) does none. The SI unit is the joule, $1\ \text{J} = 1\ \text{N}\cdot\text{m}$.
- To find the work done by each force, draw the free-body diagram, find the unknown forces with Newton's laws, then multiply by the displacement. The net work also equals $\mathbf{F}_{\text{net}}\cdot\mathbf{d}$, and is zero whenever the velocity is constant. Run backward, $d = \Delta E/F_{\text{net}}$ gives the displacement needed for a given energy change.
- The **work done by gravity** near Earth's surface is $W_g = -mg\ \Delta y$. It depends only on the change in height, not on the path.
- For a **variable force**, $W = \int F\ dx$, the area under the force–position graph. A spring obeys Hooke's law $F_s = -kx$; stretching it from rest takes work $\frac{1}{2}kx^2$, and the spring itself does $W_s = \frac{1}{2}kx_i^2 - \frac{1}{2}kx_f^2$.
- **Power** is the rate of doing work: $P_{\text{avg}} = W/\Delta t$, $P = dW/dt = \mathbf{F}\cdot\mathbf{v}$, measured in watts. It answers "how long?" ($\Delta t = W/P$) and "how fast?" ($v = P/F$).

Key Equations

Work and energy

$$W_{\text{net}} = W_1 + W_2 + W_3 + \cdots = \Delta E \tag{7.S.1}$$

Work done by a constant force

$$W = Fd\cos\theta = \mathbf{F}\cdot\mathbf{d} \tag{7.S.2}$$

Work done by gravity near Earth's surface

$$W_g = -mg\,\Delta y = -mg\,(y_f - y_i) \tag{7.S.3}$$

Hooke's law

$$F_s = -kx \tag{7.S.4}$$

Work done by a variable force

$$W = \int_{x_i}^{x_f} F(x)\,dx \tag{7.S.5}$$

Work done by a spring

$$W_s = \tfrac{1}{2}kx_i^2 - \tfrac{1}{2}kx_f^2 \tag{7.S.6}$$

Power

$$P_{\text{avg}} = \frac{W}{\Delta t}, \qquad P = \frac{dW}{dt} = \mathbf{F}\cdot\mathbf{v} \tag{7.S.7}$$

This page titled 7.S: Work and Power (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

7.E: Work and Power (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 7.1 — Work Done by a Constant Force

7.1 ★ A worker pushes a $18\ \text{kg}$ crate $4.0\ \text{m}$ across a level floor with a horizontal force of $55\ \text{N}$.

a. How much work does the worker do?
b. If instead the force is applied at 30° below the horizontal with the same magnitude, how much work is done over the same distance?

7.2 ★ *[Conceptual.]* A satellite moves in a circular orbit at constant speed.

a. How much work does gravity do on the satellite over one full orbit? Explain in terms of the angle between force and displacement.
b. Does the satellite's energy change over the orbit? Explain.

7.3 ★★ A $25\ \text{kg}$ sled is pulled at constant velocity across level snow by a rope inclined 30° above the horizontal. The coefficient of kinetic friction is $\mu_k = 0.12$, and the sled moves $15\ \text{m}$.

a. Draw the free-body diagram for the sled.
b. Use Newton's laws to find the tension in the rope and the normal force on the sled.
c. Find the work done on the sled by each of the four forces.
d. What is the net work? Does the sled's energy change?

7.4 ★★ A $4.0\ \text{kg}$ box is pushed across a level floor by a $40\ \text{N}$ force directed 35° *below* the horizontal. The coefficient of kinetic friction is 0.25.

a. Draw the free-body diagram and find the normal force. Why is it larger than the box's weight?
b. Find the net force on the box.
c. How far must the box be pushed for its energy to increase by $60\ \text{J}$?

7.5 ★★ A $30\ \text{kg}$ box rides in the bed of a pickup truck. The truck accelerates from a stop sign at $1.5\ \text{m/s}^2$ for $40\ \text{m}$, and the box does not slip.

a. Draw the free-body diagram for the box. Which force gives it its forward acceleration?
b. Find the magnitude and direction of that force.
c. How much work does it do on the box (as seen from the ground) over the $40\ \text{m}$? Does the box's energy increase or decrease?

7.6 ★★ Two forces act on a hockey puck as it slides across frictionless ice: $\mathbf{F}_1 = (4.0\ \hat{\imath} + 3.0\ \hat{\jmath})\ \text{N}$ and $\mathbf{F}_2 = (-2.0\ \hat{\imath} + 5.0\ \hat{\jmath})\ \text{N}$. During the time they act, the puck's displacement is $\mathbf{d} = (3.0\ \hat{\imath} - 1.0\ \hat{\jmath})\ \text{m}$.

a. Find the work done by each force, using the dot product in component form.
b. What is the net work? By how much, and in which direction, does the puck's energy change?
c. Find the angle between $\mathbf{F}_1$ and $\mathbf{d}$.

7.7 ★★ *[Real-world: transportation.]* A $1200\ \text{kg}$ car traveling on a level road slams on its brakes. The wheels lock, and the car skids $25\ \text{m}$ before stopping. The coefficient of kinetic friction between the tires and the road is 0.60.

a. Draw the free-body diagram for the car during the skid, and use Newton's laws to find the normal force and the friction force.
b. Find the work done on the car by friction, by gravity, and by the normal force during the skid.
c. What is the net work? Describe what happens to the car's energy, and where that energy goes.

7.8 ★ *[Conceptual.]* A weightlifter holds a $100\ \text{kg}$ barbell motionless over her head for $5\ \text{s}$.

a. How much work does she do on the barbell during those $5\ \text{s}$? Explain using $W = Fd\cos\theta$.
b. Can the net work on an object be zero even though several of the forces on it do nonzero work? Give an example from this section.

Section 7.2 — Work Done by Gravity

7.9 ★★ A $1200\ \text{kg}$ elevator is lowered $8.0\ \text{m}$ at constant speed.

a. How much work does gravity do on the elevator?
b. How much work does the supporting cable do?
c. What is the net work?

7.10 ★★ *[Real-world: sports.]* A $7.3\ \text{kg}$ shot-put is pushed $0.55\ \text{m}$ along a straight line directed 40° above the horizontal, by a $310\ \text{N}$ force applied along that same line.

a. How much work does the athlete do on the shot during this push?
b. How much work does gravity do during the same push?

7.11 ★ *[Real-world: hiking.]* A $70\ \text{kg}$ hiker follows a switchback trail $2.5\ \text{km}$ long to the top of a ridge $450\ \text{m}$ above the trailhead.

a. How much work does gravity do on the hiker during the climb?
b. Why does the length of the trail not matter?
c. How much work does gravity do on the hiker over the round trip, back to the trailhead?

7.12 ★★ A $3.0\ \text{kg}$ block slides $2.5\ \text{m}$ down a ramp inclined at 30°. The coefficient of kinetic friction is 0.20.

a. Draw the free-body diagram and find the normal force and the friction force.
b. Find the work done on the block by gravity, the normal force, and friction.
c. What is the net work, and by how much does the block's energy change?

7.13 ★★ A $50\ \text{kg}$ crate is dragged $4.0\ \text{m}$ up a ramp inclined at 20°, at constant speed, by a rope parallel to the ramp. The coefficient of kinetic friction is 0.30.

a. Use Newton's second law to find the tension in the rope.
b. Find the work done on the crate by the rope, gravity, friction and the normal force.
c. Show that the net work is zero, and explain why it must be.

7.14 ★★ An $8.0\ \text{kg}$ crate is released from rest on a ramp inclined at 25°, and slides down it. The coefficient of kinetic friction is 0.15.

a. Find the net force on the crate along the ramp.
b. How far down the ramp has the crate slid when its energy has increased by $100\ \text{J}$?
c. Over that distance, how much work does gravity do, and how much does friction do?

7.15 ★★ A crane lifts an $800\ \text{kg}$ load $15\ \text{m}$ straight up with a constant upward acceleration of $0.60\ \text{m/s}^2$.

a. Find the tension in the cable.
b. Find the work done on the load by the cable and by gravity.
c. By how much does the load's energy change?
d. If the load were instead lifted the same height at constant speed, how much work would the cable do?

7.16 ★ A $1.5\ \text{kg}$ ball is carried from the floor up to a shelf $2.0\ \text{m}$ high by two routes: straight up, and along a ramp twice as long.

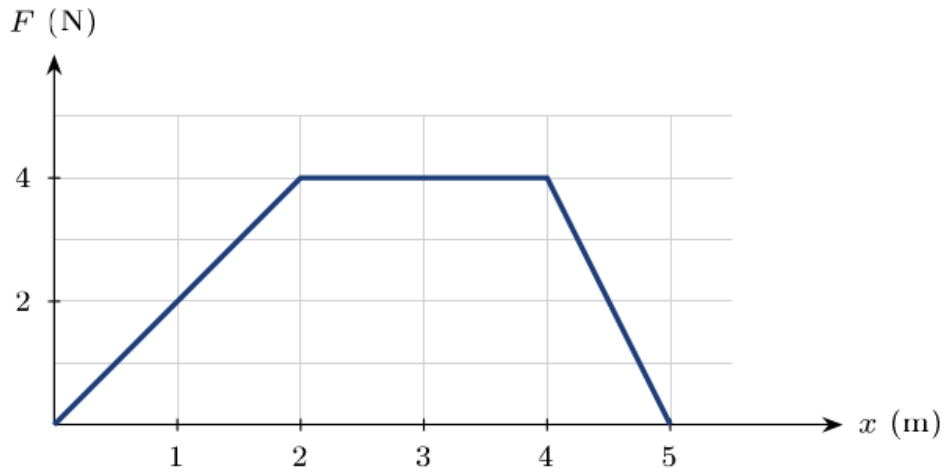
a. Find the work done by gravity for each route.
b. Explain why the two answers agree, using Figure 7.3.

Section 7.3 — Work Done by a General Force

7.17 ★ A spring has a force constant of $240\ \mathrm{N/m}$.

a. What force is needed to stretch it $0.15\ \mathrm{m}$?
b. How much work is required to stretch it that far from its relaxed length?

7.18 ★★ *[Graph-reading.]* A variable force acts on an object as it moves along the x-axis. The graph shows the force versus position.



•Figure description
A graph of a variable force in newtons versus position in meters. The force rises in a straight line from 0 at x equals 0 to 4 newtons at x equals 2 meters, stays at 4 newtons until x equals 4 meters, then falls in a straight line to 0 at x equals 5 meters. The work done is found as the area under the curve.

a. Find the work done by the force from $x = 0$ to $x = 2\ \mathrm{m}$.
b. Find the work done from $x = 0$ to $x = 5\ \mathrm{m}$ (the total area under the graph).
c. If this is the only force that does work on the object, by how much does the object's energy change between $x = 0$ and $x = 5\ \mathrm{m}$? Between $x = 4\ \mathrm{m}$ and $x = 5\ \mathrm{m}$?

7.19 ★★ *[Conceptual.]* A spring is stretched a distance x and then stretched a further equal distance x (to a total of $2x$).

a. Is the work needed for the second stretch equal to, greater than, or less than the work for the first? Justify using $\frac{1}{2}kx^2$.
b. Explain why the answer makes sense in terms of the force–position graph.

7.20 ★★ A $2.0\ \mathrm{kg}$ mass is hooked to the bottom of a vertical spring ($k = 150\ \mathrm{N/m}$) at its relaxed length, then lowered slowly by hand until it hangs at rest on its own.

a. Draw the free-body diagram for the mass when it finally hangs at rest, and use it to find how far the spring stretches.
b. Find the work done on the mass by gravity and by the spring during the lowering.
c. How much work does the hand do? What is the net work?

7.21 ★★ A block on a frictionless horizontal surface is attached to a spring ($k = 400\ \mathrm{N/m}$). The block moves from a position where the spring is compressed $0.10\ \mathrm{m}$ to one where it is stretched $0.050\ \mathrm{m}$.

a. How much work does the spring do on the block?
b. The block then moves back from the stretched position to the relaxed length. How much work does the spring do on it over this stage?
c. Over which part of the first stage does the spring do positive work, and over which part negative? Explain without calculation.

7.22 ★★ A thick rubber cord does not obey Hooke's law. Stretched a distance x from its relaxed length, it pulls back with a force of magnitude $F = ax + bx^3$, where $a = 50\ \mathrm{N/m}$ and $b = 800\ \mathrm{N/m^3}$.

a. How much work is required to stretch the cord $0.20\ \text{m}$ from its relaxed length?
b. How much would be required if the cord obeyed Hooke's law with $k = a$? Which term dominates at this stretch?
c. What force is needed to hold the cord at $x = 0.20\ \text{m}$?

Section 7.4 — Power

7.23 ★ A motor does $3600\ \text{J}$ of work in $12\ \text{s}$. What is its average power output, in watts and in horsepower?

7.24 ★★ *[Real-world: engineering.]* A pump raises $200\ \text{kg}$ of water per minute from a well $15\ \text{m}$ deep.

a. How much work is done against gravity each minute?
b. What is the minimum power the pump must supply?

7.25 ★ *[Conceptual.]* Two cranes lift identical loads to the same height, but crane A does it in half the time of crane B.

a. Which crane does more work? Which delivers more power? Explain.
b. A small motor can still raise a very heavy load given enough time. Explain how this is possible, in terms of work and power.

7.26 ★★ *[Real-world: biology.]* A $75\ \text{kg}$ student climbs a flight of stairs, rising $4.5\ \text{m}$, in $6.0\ \text{s}$ at a steady pace.

a. How much work does gravity do on the student?
b. What average power must the student's legs deliver to raise her body? Express it in watts and in horsepower.
c. How does the power change if she runs up the same stairs in $3.0\ \text{s}$?

7.27 ★★ A $900\ \text{W}$ electric winch drags a $120\ \text{kg}$ crate up a ramp inclined at 15°, pulling on a cable parallel to the ramp. The coefficient of kinetic friction is 0.25.

a. Draw the free-body diagram and find the cable tension needed to move the crate at constant speed.
b. If the winch runs at its full $900\ \text{W}$, how fast does the crate move?
c. How long does it take to drag the crate $10\ \text{m}$ up the ramp?

7.28 ★★ An elevator motor can deliver $25\ \text{kW}$. The car and its load have a combined mass of $1200\ \text{kg}$, and a constant friction force of $1500\ \text{N}$ opposes the motion.

a. What is the greatest constant speed at which the motor can raise the car?
b. At that speed, how long does it take to rise $30\ \text{m}$?

7.29 ★★ *[Real-world: transportation.]* A boat's engine delivers $50\ \text{kW}$ to the propeller while the boat cruises at a steady $12\ \text{m/s}$.

a. What is the drag force of the water on the boat? (Explain why the forward thrust equals the drag.)
b. If the drag force is proportional to v^2, what power would be needed to cruise at $15\ \text{m/s}$?

7.30 ★★★ *[Real-world: sports.]* A cyclist and bicycle of combined mass $70\ \text{kg}$ travel at a steady $8.0\ \text{m/s}$ against a total resistive force of $40\ \text{N}$.

a. What power must the cyclist supply to maintain this speed on level ground?
b. What additional power is needed to climb a $5.0\ \%$ grade (rising $5.0\ \text{m}$ per $100\ \text{m}$ traveled) at the same speed?

Challenge Problems

7.31 ★★★ A crate of mass m is dragged at constant velocity a distance d across a level floor (coefficient of kinetic friction μ_k) by a rope at an angle θ above the horizontal.

a. Use Newton's laws to show that the tension is $F_T = \dfrac{\mu_k m g}{\cos\theta + \mu_k \sin\theta}$.
b. Show that the work done by the rope is $W_T = \dfrac{\mu_k m g d}{1 + \mu_k \tan\theta}$.
c. Show that the tension is smallest when $\tan\theta = \mu_k$. For $m = 50\ \text{kg}$, $\mu_k = 0.40$ and $d = 10\ \text{m}$, find that angle, the minimum

tension, and the work done by the rope at that angle. Compare with pulling horizontally.

d. According to (b), the rope's work keeps *decreasing* as θ increases, even past the angle of least tension. Explain why, in terms of the normal force and the work done by friction.

7.32 ★★★ A single horizontal force $F(x) = (12 - 3x^2)\ \text{N}$ (with x in meters) acts on an object that starts at the origin on a frictionless surface. No other force does work on it.

a. By integrating, find the work done by F as the object moves from $x = 0$ to $x = 1.0\ \text{m}$.

b. At what position is the object's energy greatest? Explain, and find by how much its energy has increased there.

c. At what position (beyond the origin) has its energy returned to its starting value?

d. Sketch $F(x)$ for $0 \le x \le 3.5\ \text{m}$ and shade the areas that correspond to your answers to (b) and (c).

7.33 ★★★ *[Real-world: transportation.]* A loaded $20000\ \text{kg}$ truck has an engine that can deliver at most $300\ \text{kW}$ to the wheels. Rolling resistance is a constant $1960\ \text{N}$, and air drag is $F_D = \tfrac{1}{2} C \rho A v^2$ with $C = 0.70$, $\rho = 1.2\ \text{kg/m}^3$ and $A = 8.0\ \text{m}^2$.

a. Show that on level ground the top speed satisfies $(1960 + 3.36\ v^2)\ v = 3.00 \times 10^5$ in SI units, and solve it numerically.

b. Find the top speed when climbing a $5.0\ \%$ grade (rising $5.0\ \text{m}$ per $100\ \text{m}$ of road).

c. At that speed, how long does it take the truck to climb a mountain pass $600\ \text{m}$ high?

d. On the grade, what fraction of the engine's power goes into the work against gravity?

This page titled 7.E: Work and Power (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8: CONSERVATION OF ENERGY

In Chapter 7 we learned to calculate work, and we took it as a definition that work changes an object's energy. We were careful not to say what energy *is*. This chapter answers that question, and in answering it arrives at one of the deepest and most useful statements in all of physics: **energy is conserved**. It can be transformed endlessly from one form to another (a falling rock trades height for speed, a compressed spring launches a ball, friction turns motion into warmth), but the total amount never changes. Nothing in nature has ever been observed to create or destroy energy.

The concept was not part of Newton's mechanics. It took the generation of scientists after him to recognize that, beneath the endless variety of motions and forces, a single bookkeeping quantity is transferred and transformed but never lost. Energy comes in many forms (the energy of motion, the energy stored by position or configuration, heat, light, chemical energy, and more), all measured in the same unit, the joule. In mechanics two forms do almost all the work: *kinetic* energy, which an object has because it is moving, and *potential* energy, which a system has because of how its parts are arranged.

The practical payoff is enormous. With conservation of energy we can connect the start and end of a process without following the messy details in between. We need not know the forces at every instant, integrate any equation of motion, or even know the path taken: only the energy at the beginning and the end. A problem that would be a nightmare with Newton's second law often collapses to a single line of energy bookkeeping.

Learning Objectives

- Describe how work done on a system changes its kinetic energy, its potential energy, or both.
- Define and calculate kinetic energy, and apply the work–kinetic energy theorem.
- Define and calculate gravitational and elastic potential energy, and choose a reference level.
- Distinguish conservative from nonconservative forces, and relate a conservative force to its potential energy.
- Apply conservation of mechanical energy, and extend it to problems with friction.
- Read a potential-energy curve: locate turning points and classify equilibria as stable or unstable.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

8.1: Where Does the Work Go?
8.2: Kinetic Energy and the Work–Kinetic Energy Theorem
8.3: Potential Energy
8.4: Conservative and Nonconservative Forces
8.5: Conservation of Mechanical Energy
8.S: Conservation of Energy (Summary)
8.E: Conservation of Energy (Problems)

This page titled 8: Conservation of Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.1: Where Does the Work Go?

Chapter 7 ended many examples with a sentence like "the block's energy increases by 41.7 J." To make that sentence mean something we need to be precise about two things: *whose* energy, and *what kind*.

The "whose" is answered by choosing a **system**: the object, or collection of objects, whose energy we are keeping track of. Everything else is the *environment*. A force exerted on the system by something in the environment is an *external* force, and its work transfers energy between the environment and the system. That is all equation (7.1) said: the work done on a system from outside changes the system's energy.

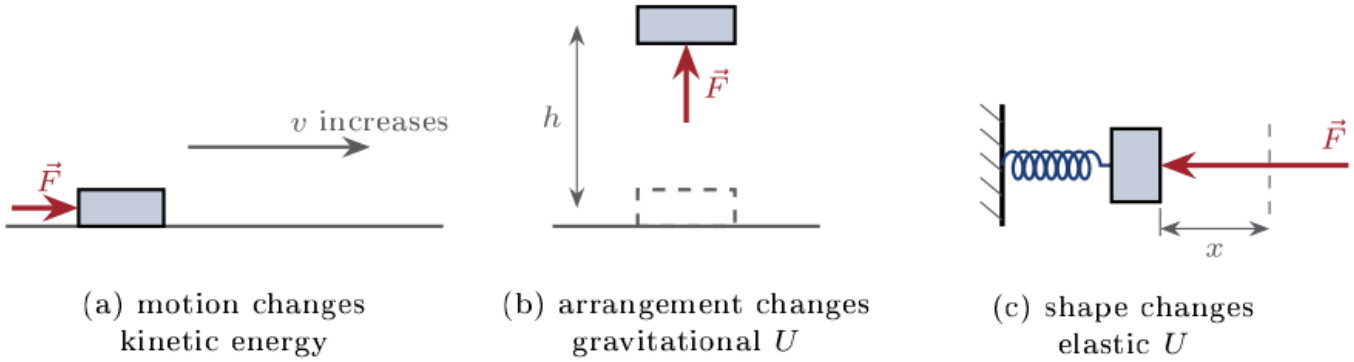


Figure 8.1. Where the work goes. In each panel a hand does positive work on a system. In (a) the work changes the system's motion; in (b) and (c) it changes the system's arrangement. The first kind of energy is kinetic, the second potential.

•Figure description

Three panels. In panel a, a hand pushes a puck across ice and the puck speeds up: the work becomes energy of motion, kinetic energy. In panel b, a hand lifts a book a height h above the ground at constant speed: the work becomes energy of arrangement of the book and Earth, gravitational potential energy. In panel c, a hand compresses a spring a distance x: the work becomes energy of the spring's shape, elastic potential energy.

The "what kind" is best answered by looking at three simple cases, each with a hand doing positive work on something (Figure 8.1).

a. **A puck pushed across frictionless ice.** The hand pushes forward as the puck moves forward, doing positive work. Nothing else does any work: gravity and the normal force are perpendicular to the motion. The puck speeds up. The work has gone into the energy of the puck's *motion*.
b. **A book lifted at constant speed.** The hand does $+mgh$ of work. If the system is the book alone, gravity (exerted by the Earth, outside the system) does $-mgh$, the net work is zero, and the book's energy does not change, exactly the bookkeeping of Chapter 7. Now take the system to be the book *and the Earth*. Gravity is an interaction between two parts of the system, and the only external force doing work is the hand. The system has gained mgh of energy, but nothing in it is moving any faster. What has changed is the *arrangement* of the system: the book and the Earth are farther apart.
c. **A spring compressed slowly.** The hand does $+\frac{1}{2}kx^2$ of work on the spring (equation (7.7)). Again nothing ends up moving; what has changed is the spring's *shape*.

These are the two basic kinds of mechanical energy. **Kinetic energy**, the energy of motion, belongs to anything that moves. **Potential energy**, the energy of arrangement, belongs to a system whose parts exert forces on one another: the book and the Earth, the two ends of a spring. The work done on a system by external forces goes into one or both:

Key Equation: Energy of a system

$$W_{\text{ext}} = \Delta E_{\text{sys}} = \Delta K + \Delta U. \tag{8.1}$$

There is a third possibility, which we take up in §8.4: when surfaces within the system rub against each other, some of the energy goes into *thermal energy*, the random jiggling of atoms that we feel as warmth, and a term ΔE_{th} joins the right-hand side.

Notice which work appears on the left of equation (8.1): only the work of *external* forces. The parts of a system also push and pull on one another, and those *internal* forces do work too. But when an internal force is conservative, like gravity between the book and the Earth or the force between the two ends of a spring, its work is exactly what the change in potential energy keeps track of (Section 8.3). Counting it again on the left would count it twice. So the bookkeeping splits cleanly: external forces do the work W_{ext}, and internal conservative forces are represented by ΔU.

Equation (8.1) is the outline of this whole chapter. Section 8.2 supplies the formula for K; Section 8.3 supplies the formulas for U ; and Section 8.5 asks what happens when *no* external work is done at all.

Checkpoint 8.1

A ball falls from a shelf to the floor. Choose the system to be (a) the ball alone, and then (b) the ball and the Earth. In each case, is the work done by gravity external work? Does the system's kinetic energy change? Its potential energy?

Answer

Answers to Checkpoints

Checkpoint 8.1. (a) With the ball alone as the system, the Earth is outside it, so gravity is an external force. It does positive work mgh, and the ball's kinetic energy increases by that amount. A single ball has no internal arrangement, so there is no potential energy to change. (b) With the ball and the Earth as the system, gravity acts between two parts of the system and is not external. No external force does work, so the system's total energy is unchanged: its kinetic energy rises by mgh as its gravitational potential energy falls by mgh. The two descriptions agree about the ball's speed; they just keep the books differently.

This page titled 8.1: Where Does the Work Go? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.2: Kinetic Energy and the Work–Kinetic Energy Theorem

Start with the simplest system: a single object treated as a particle, with no internal parts to rearrange. Every force on it is exerted by something outside it, so every force is external, and the net work W_{net}, the sum of the works done by all the forces on its free-body diagram, is the W_{ext} of equation (8.1). A particle on its own has no arrangement and so no potential energy; its energy can only be energy of motion. Equation (7.1) of Chapter 7 therefore says that the net work changes the particle's energy of motion. What we need is a formula for that energy, and Newton's second law supplies it.

Deriving the theorem

Let an object of mass m move along the x-axis under a net force F_{net}, which may change as the object moves. As it goes from x_i to x_f, its speed changes from v_i to v_f. The net work is the work integral of Chapter 7 applied to the net force:

$$W_{\text{net}} = \int_{x_i}^{x_f} F_{\text{net}}\, dx. \tag{8.2.1}$$

Newton's second law replaces F_{net} by ma. To connect the acceleration to the speed, use the chain rule:

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}x}\,\frac{\mathrm{d}x}{\mathrm{d}t} = v\,\frac{\mathrm{d}v}{\mathrm{d}x}. \tag{8.2.2}$$

Substituting, the dx's cancel and the integral becomes an integral over the speed. As x runs from x_i to x_f, v runs from v_i to v_f:

$$W_{\text{net}} = \int_{x_i}^{x_f} mv\,\frac{\mathrm{d}v}{\mathrm{d}x}\,dx = \int_{v_i}^{v_f} mv\,dv = \tfrac{1}{2}mv_f^2 - \tfrac{1}{2}mv_i^2. \tag{8.2.3}$$

Nothing here assumed the force was constant, so the result holds for any net force. A new quantity has appeared on its own: $\frac{1}{2}mv^2$, built only from the object's mass and speed. It is the energy of motion we were looking for. We call it the **kinetic energy**:

Key Equation: Kinetic energy

$$K = \tfrac{1}{2}mv^2. \tag{8.2}$$

and the result above is one of the central theorems of mechanics, the **work–kinetic energy theorem**:

Key Equation: Work–kinetic energy theorem

$$W_{\text{net}} = \Delta K = \tfrac{1}{2}mv_f^2 - \tfrac{1}{2}mv_i^2. \tag{8.3}$$

In words: the net work done on an object equals the change in its kinetic energy. Positive net work speeds an object up; negative net work slows it down; zero net work leaves its speed unchanged. This is equation (7.1) of Chapter 7 with the energy finally named: for a particle, the change in energy ΔE is the change in the kinetic energy ΔK.

The theorem is often the fastest route to a final speed, because it sidesteps time and acceleration entirely: only the net work, the mass, and the two speeds appear.

Kinetic energy

Like work, kinetic energy is measured in joules (Section 7.1): $(\text{ kg})\ (\text{m/s})^2 = \text{ kg}\cdot\text{m}^2/\text{s}^2 = \text{ J}$, a unit named for the nineteenth-century English physicist James Prescott Joule. A few features are worth noting. Kinetic energy is a *scalar*: it has no direction, and a $20\ \text{kg}$ object moving north has the same kinetic energy as one moving south at the same speed. It is never negative, since it depends on v^2. And that dependence on the *square* of the speed has real consequences: doubling an object's speed *quadruples* its kinetic energy, which is why stopping distances grow so steeply with speed, and why a highway collision is far more than twice as destructive as a parking-lot one.

Checkpoint 8.2

A car speeds up from 20 to $40\ \mathrm{km/h}$, and later from 40 to $60\ \mathrm{km/h}$. Both are $20\ \mathrm{km/h}$ increases. Does the kinetic energy rise by the same amount each time? If not, which change is larger, and by what factor?

Answer

Example 8.1: Which has more kinetic energy?

A $0.057\ \mathrm{kg}$ tennis ball is served at $50\ \mathrm{m/s}$. A $1200\ \mathrm{kg}$ car rolls through a parking lot at a brisk walking pace, $1.5\ \mathrm{m/s}$. Which has more kinetic energy, and by how much?

Solution

In $K = \frac{1}{2}mv^2$ only the speed is squared, so square it first and then multiply: squaring $\frac{1}{2}mv$ as a whole is a common slip. For the tennis ball,

$$K_{\mathrm{ball}} = \tfrac{1}{2}(0.057\,\mathrm{kg})(50\,\mathrm{m/s})^2 = \tfrac{1}{2}(0.057)(2500)\,\mathrm{J} = 71\,\mathrm{J}, \tag{8.2.4}$$

and for the car,

$$K_{\mathrm{car}} = \tfrac{1}{2}(1200\,\mathrm{kg})(1.5\,\mathrm{m/s})^2 = \tfrac{1}{2}(1200)(2.25)\,\mathrm{J} = \boxed{1.4\times10^{3}\,\mathrm{J}}, \tag{8.2.5}$$

so the slow-moving car has $1350/71 \approx 19$ times the kinetic energy of the fast-moving ball.

The comparison is worth dwelling on, because it shows mass and speed do not enter on equal footing. Write the ratio of the two energies; the factors of $\frac{1}{2}$ cancel, and the ratio splits into a mass part and a speed part:

$$\frac{K_{\mathrm{car}}}{K_{\mathrm{ball}}} = \frac{m_{\mathrm{car}}}{m_{\mathrm{ball}}} \times \left(\frac{v_{\mathrm{car}}}{v_{\mathrm{ball}}}\right)^2 = \frac{1200}{0.057} \times \left(\frac{1.5}{50}\right)^2 \approx (2.1\times10^{4})(9.0\times10^{-4}) \approx 19. \tag{8.2.6}$$

The car outweighs the ball by a factor of about 2.1×10^{4}, while the ball moves $50/1.5 \approx 33$ times faster, and since speed is *squared*, that becomes a factor of $33^2 \approx 1.1\times10^{3}$ in the ball's favor. The mass advantage still wins, but only by the ratio of those two numbers, about 19. Notice too how steeply the car's energy climbs with speed: driven at $30\ \mathrm{m/s}$, twenty times faster, the same car would carry $20^2 = 400$ times as much, $5.4\times10^{5}\ \mathrm{J}$.

Using the theorem

Two cautions are worth stating carefully, because both are common sources of wrong answers.

First, W_{net} means the work done by *every* force acting on the object, added together. Leave out gravity, or friction, or the pull of a second rope, and the theorem will give the wrong speed, not because the theorem failed, but because the wrong quantity was put into it. The free-body diagram is still the place to start: it is the list of forces whose work you must add.

Second, ΔK is the change in *kinetic* energy alone. Lift a book of mass m a height h at constant speed: your hand does $+mgh$ of work and gravity does $-mgh$, so $W_{\mathrm{net}} = 0$, and indeed $\Delta K = 0$; the book is no faster at the top. The theorem holds exactly. What it does not tell us is that the book–Earth system now holds energy it did not hold before. Tracking that stored energy is the job of potential energy, in §8.3; it is a separate account, not an exception to this one.

Example 8.2: Speed from net work

A $6.0\ \mathrm{kg}$ block, initially at rest, is pulled along a frictionless horizontal surface by a constant horizontal force of $12\ \mathrm{N}$. How fast is it moving after it has traveled $3.0\ \mathrm{m}$?

Solution

Step 1: Net work. The only force doing work is the applied force: gravity and the normal force are perpendicular to the motion, and there is no friction. The applied force is along the displacement ($\theta = 0$, $\cos\theta = 1$), so

$$W_{\text{net}} = Fd = (12\,\text{N})(3.0\,\text{m}) = 36\,\text{J}. \tag{8.2.7}$$

Step 2: Solve the work–kinetic energy theorem for v_f. With $v_i = 0$, the $\frac{1}{2}mv_i^2$ term vanishes and $W_{\text{net}} = \frac{1}{2}mv_f^2$. To isolate v_f, multiply both sides by 2, divide both sides by m, and take the square root:

$$W_{\text{net}} = \tfrac{1}{2}mv_f^2 \implies v_f^2 = \frac{2W_{\text{net}}}{m} \implies v_f = \sqrt{\frac{2W_{\text{net}}}{m}}. \tag{8.2.8}$$

Substituting,

$$v_f = \sqrt{\frac{2(36\,\text{J})}{6.0\,\text{kg}}} = \sqrt{12\,\text{m}^2/\text{s}^2} \approx \boxed{3.5\,\text{m/s}}. \tag{8.2.9}$$

The units work because $\text{J/kg} = (\text{kg}\cdot\text{m}^2/\text{s}^2)/\text{kg} = \text{m}^2/\text{s}^2$.

Check. Solve it the Chapter 2 way instead. The acceleration is $a = F/m = 12/6.0 = 2.0\ \text{m/s}^2$, and $v_f^2 = v_i^2 + 2ad = 0 + 2(2.0)(3.0) = 12\ \text{m}^2/\text{s}^2$: the same $\sqrt{12}$. The work–kinetic energy theorem got there without ever computing an acceleration.

Example 8.3: Back to Chapter 7: how fast?

Three examples in Chapter 7 ended with an object's energy increasing by a known amount. In each case the object started from rest. Find its final speed.

a. The $40\ \text{kg}$ crate dragged until its energy had increased by $500\ \text{J}$ (Example 7.3).
b. The $5.0\ \text{kg}$ block whose energy increased by $41.7\ \text{J}$ sliding $3.0\ \text{m}$ down a rough ramp (Example 7.4).
c. The $100\ \text{g}$ rifle bullet on which the gas did $9.00\ \text{kJ}$ of work.

Solution

In each case the object is a particle starting from rest, so the energy increase *is* its kinetic energy at the end: $\frac{1}{2}mv_f^2 = \Delta K$, and $v_f = \sqrt{2\,\Delta K/m}$.

(a) The crate:

$$v_f = \sqrt{\frac{2(500\,\text{J})}{40\,\text{kg}}} = \sqrt{25\,\text{m}^2/\text{s}^2} = \boxed{5.0\,\text{m/s}}. \tag{8.2.10}$$

(b) The block on the rough ramp:

$$v_f = \sqrt{\frac{2(41.7\,\text{J})}{5.0\,\text{kg}}} = \sqrt{16.7\,\text{m}^2/\text{s}^2} \approx \boxed{4.1\,\text{m/s}}. \tag{8.2.11}$$

(c) The bullet. Convert the mass to SI first, $100\ \text{g} = 0.100\ \text{kg}$:

$$v_f = \sqrt{\frac{2(9000\,\text{J})}{0.100\,\text{kg}}} = \sqrt{1.80\times10^5\,\text{m}^2/\text{s}^2} \approx \boxed{424\,\text{m/s}}. \tag{8.2.12}$$

Check on (b). Newton's second law gave a net force of $13.9\ \text{N}$ down the ramp, so $a = 13.9/5.0 = 2.78\ \text{m/s}^2$, and $v_f^2 = 2ad = 2(2.78)(3.0) = 16.7\ \text{m}^2/\text{s}^2$. ✓ The free-body diagram and Newton's laws found the forces; the work–kinetic energy theorem turned them into a speed without any kinematics.

Example 8.4: Stopping distance

A $1500\ \text{kg}$ car traveling at $25\ \text{m/s}$ on level pavement locks its brakes and skids to a stop. The coefficient of kinetic friction between tires and road is 0.70. (a) How far does the car skid? (b) How far would it skid from twice the speed?

Solution

Step 1: Free-body diagram. Three forces act: gravity mg down, the normal force F_N up, and kinetic friction $F_{f,k}$ backward. Vertically there is no acceleration, so $F_N = mg$ and $F_{f,k} = \mu_k mg$.

Step 2: Net work. Gravity and the normal force are perpendicular to the motion and do no work. Friction is opposite the displacement d, so $W_{\text{net}} = -\mu_k mg\ d$.

Step 3: The theorem. The car ends at rest, so $\Delta K = 0 - \frac{1}{2}mv_i^2$:

$$-\mu_k mg\, d = -\tfrac{1}{2}mv_i^2. \tag{8.2.13}$$

Cancel m (every term has it), cancel the minus signs, and divide both sides by $\mu_k g$:

$$d = \frac{v_i^2}{2\mu_k g} = \frac{(25\,\text{m/s})^2}{2(0.70)(9.8\,\text{m/s}^2)} = \frac{625}{13.7}\,\text{m} \approx \boxed{46\,\text{m}}. \tag{8.2.14}$$

(b) The skid distance goes as v_i^2, so doubling the speed multiplies it by four: $\boxed{1.8 \times 10^2\,\text{m}}$.

Check. The mass canceled, as it should: a heavier car has more kinetic energy to lose, but it also presses harder on the road and has proportionally more friction. And from kinematics, $a = -\mu_k g = -6.86\ \text{m/s}^2$ and $d = v_i^2/(2|a|) = 625/13.7 = 46\ \text{m}$. ✓

Answers to Checkpoints

Checkpoint 8.2. No. Kinetic energy goes as v^2, so equal steps in speed give larger and larger steps in energy. In units of $\frac{1}{2}m$ $(\text{km/h})^2$, the first change is $40^2 - 20^2 = 1600 - 400 = 1200$ and the second is $60^2 - 40^2 = 3600 - 1600 = 2000$. The second change is larger by a factor of $2000/1200 = 5/3 \approx 1.7$, and by the work–kinetic energy theorem it demands correspondingly more work from the engine.

This page titled 8.2: Kinetic Energy and the Work–Kinetic Energy Theorem is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.3: Potential Energy

Not all work ends up as motion. Lift a basketball slowly to a shelf and it is no faster at the top than at the bottom, yet you did work mgh on it. That energy is not lost; it is stored, ready to reappear as motion if the ball falls. Energy stored by virtue of the arrangement of a system is called **potential energy**, written U.

Potential energy is tied to a force acting *within* the system: gravity between the ball and the Earth, the elastic force between the two ends of a spring. In Chapter 7 we found that each of these forces does work that depends only on where the object starts and ends, not on the route between. That is what makes it possible to attach a single number U to each arrangement. We define the potential energy so that the work W_{int} done by the internal force, the force one part of the system exerts on another, is the *decrease* in U:

$$W_{\text{int}} = -\Delta U. \tag{8.4}$$

When the force does positive work (gravity on a falling ball, a spring pushing a block back toward its relaxed length) the potential energy goes down, and the energy reappears somewhere else, usually as kinetic energy. When the force does negative work, potential energy is stored. Section 8.4 explains why only some forces have a potential energy; here we work out the two that matter most.

Gravitational potential energy

By equation (7.3), gravity near the Earth's surface does work $W_g = -mg\ \Delta y = -(mgy_f - mgy_i)$. Comparing with equation (8.4), the quantity whose change this is, is mgy. That is the **gravitational potential energy**:

Key Equation: Gravitational potential energy

$$U_g = mgy, \tag{8.5}$$

where y is the height above a chosen reference level. (Many books write $U_g = mgh$, with h the height; it is the same thing.) When you lift the basketball a height h at constant speed, $\Delta U_g = +mgh$, which is exactly the mgh of work your hand did in equation (7.4): potential energy is stored work.

Only *changes* in potential energy are physically meaningful, so we are free to put the reference level ($U = 0$) wherever it is convenient: the floor, a tabletop, the bottom of a hill. This freedom takes some getting used to. It means U_g has no single "correct" value, and that it is perfectly ordinary for U_g to be *negative*: an object in a basement is below a reference level chosen at street height, so $y < 0$ and $U_g < 0$. Nothing is wrong when that happens. What matters, and what never depends on the choice, is the difference $\Delta U_g = mg\ \Delta y$ between two positions. Kinetic energy, by contrast, has no such freedom: K is never negative, and $K = 0$ has the unambiguous meaning that the object is at rest.

One further remark, easy to skip but worth knowing. Gravitational potential energy is not really stored "in" the object; it is stored in the *configuration* of the object and the Earth together, in how far apart they are. We speak of "the ball's potential energy" as a convenient shorthand, but it belongs to the ball–Earth system, as in Figure 8.1(b).

There is another way to picture this that will matter a great deal later. The Earth surrounds itself with a *gravitational field*: at every point near the Earth, a mass m feels a force $m\mathbf{g}$, so $\mathbf{g}$ is the force *per unit mass*. It is a property of the location, not of whatever object happens to be placed there. In the same spirit, the potential energy per unit mass, $U_g/m = gy$, is also a property of the location: it says how much energy each kilogram has by virtue of being at height y in the Earth's field. Both ideas return for electric forces. In Chapter 16 the electric field is defined as the force per unit charge, and in Chapter 18 the electric potential is the potential energy per unit charge. There, the energy of a system of charges can be pictured as stored in the field itself, and the ball and the Earth are the model to keep in mind.

Example 8.5: Choosing a reference level

A $1.2\ \text{kg}$ book sits on a desk $0.80\ \text{m}$ above the floor. It is lifted to a shelf $2.1\ \text{m}$ above the floor. Find its gravitational potential energy at both positions, and the change, using (a) the floor as the reference level and (b) the desktop as the reference level.

Solution

In each case, y is the height measured *from the chosen reference level*, so the first job is to find the two heights relative to that level.

(a) With $U = 0$ at the floor, the heights are $0.80\ \text{m}$ and $2.1\ \text{m}$:

$$\begin{aligned} U_i &= (1.2\,\text{kg})(9.8\,\text{m/s}^2)(0.80\,\text{m}) = 9.4\,\text{J}, \\ U_f &= (1.2\,\text{kg})(9.8\,\text{m/s}^2)(2.1\,\text{m}) = 24.7\,\text{J}, \end{aligned} \tag{8.3.1}$$

so $\Delta U = U_f - U_i = 24.7 - 9.4 = 15.3\ \text{J}$.

(b) With $U = 0$ at the desktop, every height is measured from $0.80\ \text{m}$ up, so subtract 0.80 from each floor height. The book starts at $y = 0.80 - 0.80 = 0$ and ends at $y = 2.1 - 0.80 = 1.3\ \text{m}$:

$$U_i = 0, \qquad U_f = (1.2)(9.8)(1.3) = 15.3\,\text{J}, \tag{8.3.2}$$

so $\Delta U = 15.3 - 0 = \boxed{15.3\,\text{J}}$.

Why the change cannot depend on the choice. Moving the reference level subtracts the same amount from both heights, so it subtracts the same amount of energy from both U_i and U_f, and that common amount drops out of the difference. Algebraically, $\Delta U = mgy_f - mgy_i = mg(y_f - y_i) = mg\ \Delta y$, and $\Delta y = 2.1 - 0.80 = 1.3\ \text{m}$ whichever level we measure from. So $\Delta U = (1.2)(9.8)(1.3) = 15.3\ \text{J}$ in one step.

The individual values disagree ($9.4\ \text{J}$ versus $0\ \text{J}$ for the same book on the same desk) but the *change* is $15.3\ \text{J}$ either way, and it is the change that physics can measure. Had we placed the reference at the shelf, both values would have come out negative ($-15.3\ \text{J}$ and 0) and the change would still have been $0 - (-15.3) = 15.3\ \text{J}$. Choose whichever level makes the arithmetic easiest, state your choice, and then use it consistently for the whole problem.

Elastic potential energy

The same reasoning applies to a spring. By equation (7.8), the spring does work $W_s = \frac{1}{2}kx_i^2 - \frac{1}{2}kx_f^2 = -\left(\frac{1}{2}kx_f^2 - \frac{1}{2}kx_i^2\right)$ as its end moves from x_i to x_f. Comparing with equation (8.4), the **elastic potential energy** is

Key Equation: Elastic potential energy

$$U_s = \tfrac{1}{2}kx^2, \tag{8.6}$$

where x is the stretch or compression measured from the relaxed length. Here there is a natural reference: $U_s = 0$ when the spring is relaxed, and U_s is positive whether the spring is stretched or compressed. And again, U_s is exactly the work $\frac{1}{2}kx^2$ needed to stretch the spring in the first place (equation (7.7)). The energy is in the stretched spring, not in the block pressed against it.

Checkpoint 8.3

A spring is stretched $2.0\ \text{cm}$ from its relaxed length, and then compressed $2.0\ \text{cm}$ from its relaxed length. In which state does it store more elastic potential energy? What happens to U_s if the stretch is tripled?

Answer

Example 8.6: A spring-loaded dart

A toy dart gun has a spring with $k = 850\ \mathrm{N/m}$, compressed $6.0\ \mathrm{cm}$ when cocked. (a) How much elastic potential energy is stored? (b) The $25\ \mathrm{g}$ dart leaves the spring as it reaches its relaxed length. Treating the barrel as frictionless and horizontal, how fast is the dart moving?

Solution

(a) Convert to SI first, $x = 0.060\ \mathrm{m}$, then square before multiplying:

$$U_s = \tfrac{1}{2}kx^2 = \tfrac{1}{2}(850\,\mathrm{N/m})(0.060\,\mathrm{m})^2 = \tfrac{1}{2}(850)(0.0036)\,\mathrm{J} = \boxed{1.53\,\mathrm{J}}. \tag{8.3.3}$$

(b) As the spring relaxes from $x_i = 0.060\ \mathrm{m}$ to $x_f = 0$, it does work $W_s = \tfrac{1}{2}kx_i^2 - 0 = 1.53\ \mathrm{J}$ on the dart: equal to the drop in its potential energy. No other force does work (the barrel is frictionless and level), so by the work–kinetic energy theorem, with the dart starting from rest,

$$\tfrac{1}{2}mv^2 = 1.53\,\mathrm{J} \implies v = \sqrt{\frac{2(1.53\,\mathrm{J})}{0.025\,\mathrm{kg}}} = \sqrt{122\,\mathrm{m^2/s^2}} \approx \boxed{11\,\mathrm{m/s}}. \tag{8.3.4}$$

The stored potential energy has become kinetic energy, joule for joule. That observation, turned into a general law, is the subject of §8.5.

Answers to Checkpoints

Checkpoint 8.3. The same in both. $U_s = \tfrac{1}{2}kx^2$ depends on x^2, so a stretch and a compression of equal size store equal energy. Tripling the stretch multiplies U_s by $3^2 = 9$.

This page titled 8.3: Potential Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.4: Conservative and Nonconservative Forces

Why do gravity and springs have a potential energy, while a force like friction does not? The forces of mechanics fall into two camps, distinguished by what happens to the energy they handle.

A **conservative force** stores energy reversibly: any work it removes from an object's motion is banked as potential energy and can be fully recovered later. Gravity and the spring force are the prime examples. Lift a ball and the work goes into gravitational potential energy; let it fall and you get every joule back as kinetic energy. The defining technical property is the one we noticed in Chapter 7: the work done by a conservative force depends only on the start and end points, not on the path between them (Figure 7.3). It follows that the work it does around any closed loop (any path that returns to its starting point) is exactly zero.

A **nonconservative force** does not store energy reversibly; it transfers mechanical energy irretrievably into other forms, usually thermal energy. Kinetic friction is the archetype. Drag a crate in a circle back to where it started and friction has done negative work the whole way: the energy it removed is gone as warmth in the crate and the floor, not recoverable by returning to the start. The work done by a nonconservative force *does* depend on the path (a longer path means more friction), which is the mirror image of the conservative case.

Example 8.7: Two routes across a floor

A $20\ \text{kg}$ crate is slid across a level floor ($\mu_k = 0.30$) from one corner of a room to a point $3.0\ \text{m}$ east and $4.0\ \text{m}$ north of it, by two routes: straight along the diagonal, and east then north along the walls. Compare the work done by friction and by gravity along the two routes.

Solution

The diagonal route is $\sqrt{3.0^2 + 4.0^2} = 5.0\ \text{m}$ long; the route along the walls is $3.0 + 4.0 = 7.0\ \text{m}$. On level ground $F_N = mg$, so friction has magnitude $\mu_k mg = (0.30)(20)(9.8) = 58.8\ \text{N}$, and on every segment it points opposite the motion. Its work is minus that force times the *path length*:

$$W_{f,\text{diagonal}} = -(58.8)(5.0) = \boxed{-294\,\text{J}}, \qquad W_{f,\text{walls}} = -(58.8)(7.0) = \boxed{-412\,\text{J}}. \tag{8.4.1}$$

Gravity does $W_g = -mg\ \Delta y = 0$ along either route, since neither changes the height. Friction's work depends on the route; gravity's does not. That is the whole difference between a nonconservative and a conservative force, in one example.

This distinction is what makes potential energy possible: only a conservative force has a well-defined potential energy, because only then is the "stored" energy independent of how the object got there. Once the potential energy is defined, the force's work and the potential energy are two views of the same thing, related by the minus sign of equation (8.4). Work done *by* the force decreases the potential energy, and work done *against* it increases the potential energy:

Key Equation: Work done by a conservative force

$$W_{\text{cons}} = -\Delta U = U_i - U_f. \tag{8.7}$$

Check it against gravity. A ball falling a height h has gravity doing $W = +mgh$ on it, while its gravitational potential energy drops by $\Delta U = -mgh$. The signs agree. Equation (8.7) is what makes potential energy *useful*: it lets us stop tracking a conservative force's work moment by moment and simply carry a U term instead.

Note: Never count the same force twice

Once you write a $U_g = mgy$ term into a problem, gravity's work is *already in the books*. Adding a separate W_{gravity} alongside it counts the same energy twice: this is the single most common error in this chapter. The same goes for a spring: use either $U_s = \frac{1}{2}kx^2$ or the spring's work, never both.

The rule to hold onto: every force acting on the system enters the bookkeeping exactly *once*: as a potential-energy term if it is conservative, or as a work term if it is not.

Checkpoint 8.4

You slide a book across a table; it coasts a short way and stops. Friction has clearly removed the book's kinetic energy. Has energy been destroyed? If not, where did it go, and in what sense is energy still conserved?

Answer

Force from potential energy

So far potential energy has been something we compute *from* a force. The relationship runs the other way as well. Consider a conservative force F_x acting on a particle that moves a small distance dx along the x-axis. The work it does is $dW = F_x\ dx$, and by equation (8.7) that work is $-dU$. Setting the two equal and dividing by dx:

Key Equation: Force from a potential-energy function

$$F_x = -\frac{\mathrm{d}U}{\mathrm{d}x}. \tag{8.8}$$

The force is the negative slope of the potential-energy curve. In words that are worth memorizing: **the force always points "downhill" on the U curve.** Where U rises steeply to the right, the force is large and points left; where the curve is flat, there is no force at all. We will put this to work in §8.5.

Example 8.8: Recovering two forces you already know

Use equation (8.8) to find the force from (a) the gravitational potential energy $U = mgy$ near Earth's surface, and (b) the elastic potential energy $U = \frac{1}{2}kx^2$ of a spring.

Solution

(a) Differentiating with respect to the vertical coordinate: m and g are constants, so they come out in front, and $\mathrm{d}y/\mathrm{d}y = 1$:

$$F_y = -\frac{\mathrm{d}}{\mathrm{d}y}(mgy) = -mg\,\frac{\mathrm{d}y}{\mathrm{d}y} = \boxed{-mg}. \tag{8.4.2}$$

A constant force of magnitude mg, and the minus sign says it points in the $-y$ direction: downward. Exactly right.

(b) For the spring, the constant $\frac{1}{2}k$ comes out in front, and the power rule gives $\mathrm{d}(x^2)/\mathrm{d}x = 2x$. The 2 from the power rule cancels the $\frac{1}{2}$:

$$F_x = -\frac{\mathrm{d}}{\mathrm{d}x}\left(\tfrac{1}{2}kx^2\right) = -\tfrac{1}{2}k\,(2x) = \boxed{-kx}. \tag{8.4.3}$$

This is Hooke's law, and again the minus sign carries the physics: the force opposes the displacement, always pushing the mass back toward $x = 0$.

Neither result is new. That is the point: equation (8.8) is not an extra law of physics, just the bookkeeping of equation (8.7) read backwards.

Answers to Checkpoints

Checkpoint 8.4. No energy has been destroyed. Friction has converted the book's kinetic energy into thermal energy: the book and the tabletop are very slightly warmer where they rubbed, and a tiny amount went into sound. *Mechanical* energy (kinetic plus

potential) has decreased, but the *total* energy, including this thermal energy, is exactly the same as before. Friction does not destroy energy; it moves it into a form we can no longer easily recover as motion.

This page titled 8.4: Conservative and Nonconservative Forces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.5: Conservation of Mechanical Energy

We can now put the pieces together. Divide the forces acting on an object into conservative ones, which have potential energies, and nonconservative ones, which do not. The net work is the sum of their works, and by the work–kinetic energy theorem it equals the change in kinetic energy:

$$W_{\text{cons}} + W_{\text{nc}} = \Delta K. \tag{8.5.1}$$

Replace W_{cons} by $-\Delta U$ (equation (8.7)) and move it to the right-hand side:

$$W_{\text{nc}} = \Delta K + \Delta U. \tag{8.5.2}$$

The sum of kinetic and potential energy is the **mechanical energy** $E_{\text{mech}} = K + U$, so this says that the mechanical energy changes by exactly the work done by the nonconservative forces:

Key Equation: Change in mechanical energy

$$\Delta E_{\text{mech}} = W_{\text{nc}}. \tag{8.9}$$

When no nonconservative force does work (no friction, no drag, no push from outside), $W_{\text{nc}} = 0$ and the mechanical energy stays constant:

Key Equation: Conservation of mechanical energy

$$K_i + U_i = K_f + U_f \qquad (\text{conservative forces only}). \tag{8.10}$$

This single equation links any two moments in a process. Kinetic energy and potential energy can trade back and forth as freely as they like, but their sum cannot change. When a nonconservative force such as friction does act, we keep its work in the books:

Key Equation: Energy equation with nonconservative forces

$$K_i + U_i + W_{\text{nc}} = K_f + U_f. \tag{8.11}$$

Equation (8.11) is the workhorse of this chapter; equation (8.10) is just the special case $W_{\text{nc}} = 0$. For friction W_{nc} is negative, so mechanical energy decreases, but the *total* energy, counting the thermal energy produced, is still conserved. That is the full statement of the conservation of energy, and it has never been found to fail.

Note: Strategy: solving a problem with energy

1. **Pick the two moments.** Energy methods connect a “before” and an “after,” nothing in between. Choose the two instants where you know the most and want the most.
2. **Choose the $U = 0$ level.** Only *changes* in potential energy matter, so the reference is yours to place. Put it at the lowest point of the motion and every height in the problem comes out positive.
3. **Sort the forces.** Each force is either conservative (it gets a U term) or nonconservative (it gets a W_{nc} term). Forces perpendicular to the motion (a normal force on a fixed surface, a rope’s tension on a pendulum) do no work and get neither. A quick free-body diagram is the safest way to be sure you have them all.
4. **Write equation (8.11) and solve.** Include a term only if it is nonzero at that moment: an object at rest has $K = 0$; an object at the reference level has $U_g = 0$; a relaxed spring has $U_s = 0$.
5. **Check.** Did the mass cancel? Should it have? Is the answer’s sign and size plausible?

Note: Why energy methods win

Consider a block released from rest at the top of a smooth, curved ramp of height h, and ask for its speed at the bottom.

With Newton's second law you are in trouble immediately. The slope changes at every point, so the direction of the normal force changes, so the acceleration is neither constant nor even in a fixed direction. There is no single free-body diagram for the whole trip, and the constant-acceleration kinematics of Chapter 2 simply do not apply. You would have to set up and integrate along the curve, and to do that you would need the curve's equation, which the problem never gave you.

With energy the whole thing is one line. The normal force is perpendicular to the motion everywhere, so it does no work; only gravity does. Then $mgh = \frac{1}{2}mv^2$, and $v = \sqrt{2gh}$.

Notice what dropped out: the shape of the ramp, the mass of the block, and the time taken. That is the trade. Energy methods will not tell you *when* the block arrives or what path it took, but when the question is only about speeds, heights, and distances, they get there in a fraction of the work.

When only conservative forces act

Example 8.9: A climbing glider

A paper glider is moving at $18\ \mathrm{m/s}$ when it is $2.0\ \mathrm{m}$ above the ground. Gliding up to an altitude of $5.0\ \mathrm{m}$ (and neglecting air resistance), how fast is it then moving?

Solution

Step 1: Is mechanical energy conserved? Only gravity does work (air resistance is neglected), so yes: $K_i + U_i = K_f + U_f$, or

$$\tfrac{1}{2}mv_i^2 + mgy_i = \tfrac{1}{2}mv_f^2 + mgy_f. \tag{8.5.3}$$

Step 2: Cancel the mass and solve for v_f. Every term contains m, so divide the whole equation by m. Then isolate the v_f term: subtract gy_f from both sides and multiply everything by 2:

$$\tfrac{1}{2}v_i^2 + gy_i = \tfrac{1}{2}v_f^2 + gy_f \implies v_f^2 = v_i^2 + 2gy_i - 2gy_f = v_i^2 + 2g(y_i - y_f). \tag{8.5.4}$$

Take the square root of both sides:

$$v_f = \sqrt{v_i^2 + 2g(y_i - y_f)}. \tag{8.5.5}$$

Step 3: Substitute, watching the sign. The glider climbs, so $y_i - y_f = 2.0 - 5.0 = -3.0\ \mathrm{m}$ is negative, and the height term *subtracts* from v_i^2:

$$v_f = \sqrt{(18)^2 + 2(9.8)(-3.0)}\,\mathrm{m/s} = \sqrt{324 - 58.8}\,\mathrm{m/s} = \sqrt{265.2}\,\mathrm{m/s} \approx \boxed{16.3\,\mathrm{m/s}}. \tag{8.5.6}$$

The glider slows as it climbs, trading kinetic energy for potential energy, and the mass never entered the answer. Had we mixed up the order and used $y_f - y_i = +3.0\ \mathrm{m}$, the glider would have come out *faster* after climbing, which is a sure sign of a sign error.

Example 8.10: A ball thrown from a building

A ball is thrown from the top of a building. When it is $4.7\ \mathrm{m}$ above the ground it is moving at $12\ \mathrm{m/s}$; later it is moving at $6.2\ \mathrm{m/s}$. At what height is it then? (Neglect air resistance.)

Solution

Step 1: Set up the energy balance. Mechanical energy is conserved. As in the previous example, divide every term by m:

$$\tfrac{1}{2}v_i^2 + gy_i = \tfrac{1}{2}v_f^2 + gy_f. \tag{8.5.7}$$

Step 2: Solve for y_f. Subtract $\frac{1}{2}v_f^2$ from both sides, then divide both sides by g:

$$gy_f = gy_i + \tfrac{1}{2}v_i^2 - \tfrac{1}{2}v_f^2 \implies y_f = y_i + \frac{v_i^2 - v_f^2}{2g}. \tag{8.5.8}$$

(Dividing $\frac{1}{2}(v_i^2 - v_f^2)$ by g gives $(v_i^2 - v_f^2)/(2g)$.)

Step 3: Substitute. Square each speed before subtracting:

$$y_f = 4.7\,\text{m} + \frac{(12)^2 - (6.2)^2}{2(9.8)}\,\text{m} = 4.7\,\text{m} + \frac{144 - 38.4}{19.6}\,\text{m} = 4.7\,\text{m} + 5.4\,\text{m} \approx \boxed{10.1\,\text{m}}. \tag{8.5.9}$$

A tempting slip is to subtract the speeds first, $(12 - 6.2)^2 = 33.6$, which gives a height gain of only $1.7\ \text{m}$. Kinetic energy depends on the *square* of the speed, and $12^2 - 6.2^2$ is not $(12 - 6.2)^2$.

The ball is higher when it is slower: consistent with energy trading from kinetic to potential. Note the method does not care whether the ball is rising or falling, or what path it took.

Example 8.11: A roller-coaster hill

A $1500\ \text{kg}$ roller-coaster car is moving at $36\ \text{m/s}$ at the bottom of a hill and slows to $3.0\ \text{m/s}$ at the top. Neglecting friction, how high is the hill?

Solution

Take the bottom of the hill as $U = 0$, so the car starts with only kinetic energy and ends with both kinetic and potential energy. Conservation of mechanical energy:

$$\tfrac{1}{2}mv_i^2 = \tfrac{1}{2}mv_f^2 + mgh. \tag{8.5.10}$$

Subtract $\frac{1}{2}mv_f^2$ from both sides and divide both sides by mg; the mass cancels:

$$mgh = \tfrac{1}{2}m\left(v_i^2 - v_f^2\right) \implies h = \frac{v_i^2 - v_f^2}{2g}, \tag{8.5.11}$$

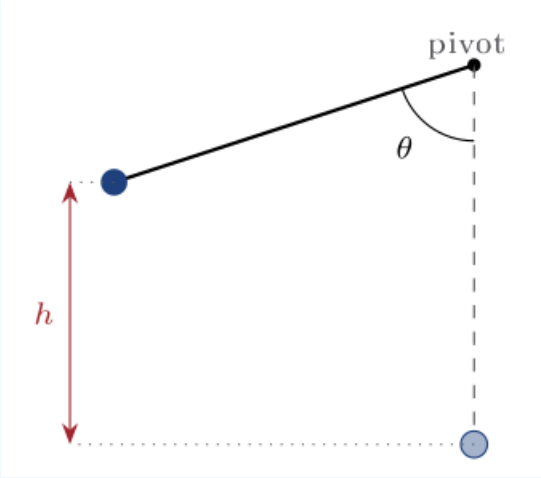
so

$$h = \frac{(36)^2 - (3.0)^2}{2(9.8)}\,\text{m} = \frac{1296 - 9}{19.6}\,\text{m} = \frac{1287}{19.6}\,\text{m} \approx \boxed{66\,\text{m}}. \tag{8.5.12}$$

The car's mass cancels: the hill height depends only on the two speeds. Notice how little the final speed matters: $(3.0)^2 = 9$ is under 1% of $(36)^2 = 1296$, so a car that barely crawls over the top has used almost all of its kinetic energy.

Example 8.12: The speed of a swing

A child swings on a $3.0\ \text{m}$ rope. If the rope makes a maximum angle of 72° with the vertical, how fast is the child moving at the lowest point?



•Figure description
A child on a swing, with the rope shown at its maximum angle theta of 72 degrees from the vertical and the arc traced down to the lowest point of the swing.

Solution

Step 1: Which forces do work? Two forces act on the child: gravity and the tension in the rope. The tension is always directed along the rope, toward the pivot, while the motion is always along the arc: perpendicular to it. A force perpendicular to the motion does no work, so the tension never appears in the energy bookkeeping, no matter how large it gets. Only gravity does work, and mechanical energy is conserved.

Step 2: Find the height drop. At the highest point the child is momentarily at rest; at the lowest point all that potential energy has become kinetic. We need the vertical distance between those two points. When the rope hangs at angle θ from the vertical, the child is a vertical distance $L\cos\theta$ below the pivot (the side of the right triangle adjacent to θ). At the bottom the child is the full length L below the pivot. The difference is the height drop:

$$h = L - L\cos\theta = L(1-\cos\theta) = (3.0\,\mathrm{m})(1-\cos 72^\circ) = (3.0\,\mathrm{m})(1-0.309) = 2.07\,\mathrm{m}. \tag{8.5.13}$$

Step 3: Energy conservation. Potential energy at the top becomes kinetic energy at the bottom. Divide both sides by m, multiply by 2, and take the square root:

$$mgh = \tfrac{1}{2}mv^2 \implies v^2 = 2gh \implies v = \sqrt{2gh} = \sqrt{2(9.8)(2.07)}\,\mathrm{m/s} \approx \boxed{6.4\,\mathrm{m/s}}. \tag{8.5.14}$$

Check. The limits of $h = L(1-\cos\theta)$ make sense: at $\theta = 0$, $\cos 0 = 1$ and $h = 0$ (no swing, no drop); at $\theta = 90^\circ$, $\cos 90^\circ = 0$ and $h = L$ (the rope horizontal, the child a full rope-length above the bottom). Our 72° gives $h = 0.69L$, between the two.

Checkpoint 8.5

Two identical marbles are released from rest at the same height on two frictionless tracks: one a straight ramp, the other a wiggly track of about the same length that dips well below the finish line before rising back to it. Which marble is moving faster at the finish? Which arrives first? Explain why these are different questions.

Answer

When nonconservative forces act

When a nonconservative force acts, mechanical energy is no longer constant; equation (8.11) accounts for the difference. The nonconservative force we meet most often is kinetic friction. A constant friction force $F_{f,k}$ opposes the motion at every point of the trip, so its work is negative and grows in proportion to the distance actually traveled:

Key Equation: Work done by kinetic friction

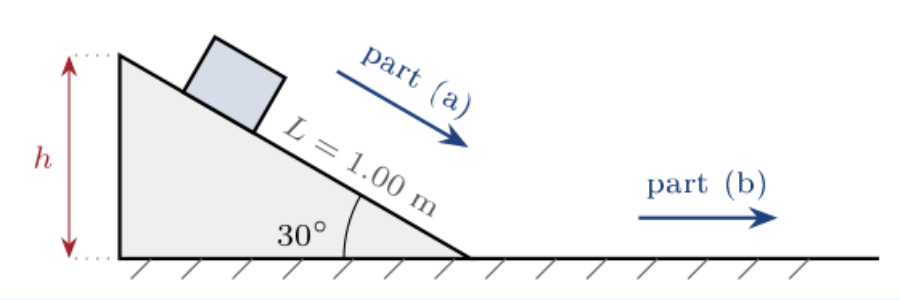
$$W_{\text{nc}} = -F_{f,k}d, \qquad F_{f,k} = \mu_k F_N. \tag{8.12}$$

The d in equation (8.12) is the **path length** (the total distance the two surfaces slide across each other) and *not* the displacement. This is precisely where the two kinds of force part ways, so it is worth dwelling on. Push a block one meter out and one meter back to where it started: the displacement is zero, so gravity and a spring do zero net work. But the surfaces rubbed for two meters, so friction has done $-2F_{f,k}$ joules of work. Whenever an object doubles back, stop and ask which length each term in your equation needs.

Note also that F_N is the normal force at the surface in question, which is *not* always mg. As in Chapters 6 and 7, it comes from the free-body diagram: on an incline at angle θ it is $mg\cos\theta$; on a horizontal floor it is mg; and if something presses the object against the surface, it is larger still.

Example 8.13: A crate down a rough ramp

A $3.00\ \text{kg}$ crate starts from rest and slides down a ramp $1.00\ \text{m}$ long, inclined at 30.0°, against a constant friction force of $5.00\ \text{N}$. (a) Find its speed at the bottom. (b) It then slides onto a horizontal floor where the friction force is also $5.00\ \text{N}$; how far does it travel before stopping? (The normal force is larger on the floor than on the ramp, so a slightly smoother floor gives the same friction force: here we are simply told the force, and need not work back to μ_k.)



•Figure description
A crate at the top of a ramp 1.00 meter long, inclined at 30 degrees, with height h. An arrow labeled part (a) points down the ramp, and an arrow labeled part (b) points along the horizontal floor beyond the foot of the ramp.

Solution

(a) Speed at the bottom. Take the bottom of the ramp as $U = 0$. The crate slides $L = 1.00\ \text{m}$ along a slope of 30.0°; the vertical drop is the side of that right triangle opposite the angle, so

$$h = L\sin 30.0^\circ = (1.00\,\text{m})(0.500) = 0.500\,\text{m}. \tag{8.5.15}$$

Friction is a nonconservative force. It points up the ramp while the crate moves down, so $\theta = 180^\circ$ and its work over the ramp is negative:

$$W_{\text{nc}} = F_f L\cos 180^\circ = -(5.00\,\text{N})(1.00\,\text{m}) = -5.00\,\text{J}. \tag{8.5.16}$$

The energy bookkeeping from top (at rest, so $K_i = 0$) to bottom (so $U_f = 0$) is $K_i + U_i + W_{\text{nc}} = K_f + U_f$, which here reduces to

$$\underbrace{mgh}_{U_i} + \underbrace{W_{\text{nc}}}_{\text{friction}} = \underbrace{\tfrac{1}{2}mv^2}_{K_f}. \tag{8.5.17}$$

With $mgh = (3.00\ \text{kg})(9.8\ \text{m/s}^2)(0.500\ \text{m}) = 14.7\ \text{J}$, and remembering that W_{nc} is negative,

$$\tfrac{1}{2}(3.00\,\text{kg})v^2 = 14.7\,\text{J} + (-5.00\,\text{J}) = 9.7\,\text{J}. \tag{8.5.18}$$

Multiply both sides by 2, divide by 3.00, and take the square root:

$$v = \sqrt{\frac{2(9.7\,\text{J})}{3.00\,\text{kg}}} = \sqrt{6.47\,\text{m}^2/\text{s}^2} \approx \boxed{2.54\,\text{m/s}}. \tag{8.5.19}$$

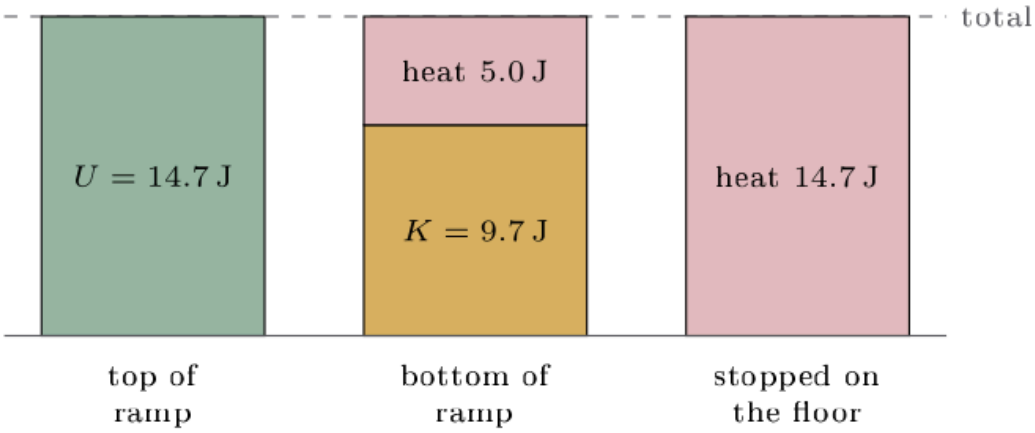
(b) Sliding distance on the floor. On the level floor there is no change in height, so no potential energy is involved. The crate's kinetic energy (9.7 J) is removed entirely by friction over a distance d; setting the energy lost equal to the work friction does, and dividing both sides by F_f,

$$\tfrac{1}{2}mv^2 = F_f d \implies d = \frac{\frac{1}{2}mv^2}{F_f} = \frac{9.7\,\text{J}}{5.00\,\text{N}} \approx \boxed{1.94\,\text{m}}. \tag{8.5.20}$$

(A joule is a newton-meter, so $\text{J}/\text{ N} = \text{ m}$.)

Check. Treat the whole trip at once. The crate starts and ends at rest, so all 14.7 J of its potential energy must go to friction, which acts over the total path $L + d$: $mgh = F_f(L + d)$, so $L + d = 14.7/5.00 = 2.94$ m and $d = 2.94 - 1.00 = 1.94$ m. ✓ The speed at the bottom was never needed for part (b).

It helps to see the whole trip as one energy budget. The crate begins with 14.7 J of gravitational potential energy and nothing else; it ends with all 14.7 J turned into thermal energy in the surfaces. Mechanical energy is *not* conserved here, but the total is, at every stage:



•Figure description
A stacked bar chart tracking energy through several stages. Each bar is divided into potential energy, kinetic energy, and energy lost to heat. The individual shares change from stage to stage while the total height stays the same.

Read the bars left to right: the potential column empties, part of it filling the kinetic column while the rest leaks away as heat; then friction on the floor drains the kinetic column too. The dashed line (the total) never moves.

Example 8.14: A block compressing a spring

A 0.80 kg block slides at 1.2 m/s toward a spring of force constant 50 N/m. (a) If the surface is frictionless, what is the maximum compression of the spring? (b) If instead there is kinetic friction with $\mu_k = 0.50$, what is the maximum compression?

Solution

At maximum compression the block is momentarily at rest, so all of its kinetic energy has gone into the spring (and, in part (b), into friction).

(a) Frictionless: $\frac{1}{2}mv^2 = \frac{1}{2}kx^2$. The factors of $\frac{1}{2}$ cancel; divide both sides by k and take the square root:

$$x^2 = \frac{mv^2}{k} \implies x = \sqrt{\frac{mv^2}{k}} = v\sqrt{\frac{m}{k}} = (1.2\,\text{m/s})\sqrt{\frac{0.80\,\text{kg}}{50\,\text{N/m}}} \approx \boxed{0.15\,\text{m}}. \tag{8.5.21}$$

(Since v is positive, $\sqrt{v^2} = v$ comes out of the root.) The units under the root are $\text{kg}/(\text{N/m}) = \text{kg}\cdot\text{m}/(\text{kg}\cdot\text{m/s}^2) = \text{s}^2$, so x comes out in $(\text{m/s})(\text{s}) = \text{m}$.

(b) Now friction also removes energy, doing work $-F_{f,k}x = -\mu_k mg\,x$ over the compression distance x (on the level floor, $F_N = mg$). The block's initial kinetic energy is shared between the spring and friction:

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}kx^2 + \mu_k mg\,x. \tag{8.5.22}$$

Compute each coefficient: $\tfrac{1}{2}mv^2 = \tfrac{1}{2}(0.80)(1.2)^2 = 0.576$ J, $\tfrac{1}{2}k = 25$ N/m, and $\mu_k mg = (0.50)(0.80)(9.8) = 3.92$ N. So

$$0.576 = 25x^2 + 3.92x. \tag{8.5.23}$$

The unknown x appears both squared and to the first power, so this is a quadratic. Subtract 0.576 from both sides to put it in standard form, and read off $a = 25$, $b = 3.92$, $c = -0.576$:

$$25x^2 + 3.92x - 0.576 = 0. \tag{8.5.24}$$

In the quadratic formula, $-4ac = -4(25)(-0.576) = +57.6$ is positive because c is negative:

$$x = \frac{-3.92 \pm \sqrt{(3.92)^2 + 57.6}}{2(25)} = \frac{-3.92 \pm \sqrt{15.4 + 57.6}}{50} = \frac{-3.92 \pm 8.54}{50}. \tag{8.5.25}$$

The "minus" root is negative, a stretch rather than a compression, and does not describe this situation. The "plus" root gives

$$x = \frac{4.62}{50} \approx \boxed{0.092\ \text{m}}. \tag{8.5.26}$$

The compression is smaller than in part (a): friction removes some energy that would otherwise have gone into the spring.

Check. At $x = 0.092$ m the spring holds $\tfrac{1}{2}(50)(0.092)^2 = 0.212$ J and friction has taken $(3.92)(0.092) = 0.361$ J. Together, $0.212 + 0.361 = 0.573$ J, which matches the initial 0.576 J to rounding. Friction took the larger share.

Reading a potential-energy curve

When mechanical energy is conserved, a single graph tells the whole story of a particle's motion along a line. A plot of $U(x)$, with the particle's total energy E drawn as a horizontal line across it, is called an **energy diagram**, and it is a remarkably efficient picture. Because $E = K + U$ and E is a constant, the vertical distance between the line and the curve *is* the kinetic energy:

$$K(x) = E - U(x). \tag{8.5.27}$$

And by equation (8.8), the slope of the curve gives the force: the force always points downhill on the U curve.

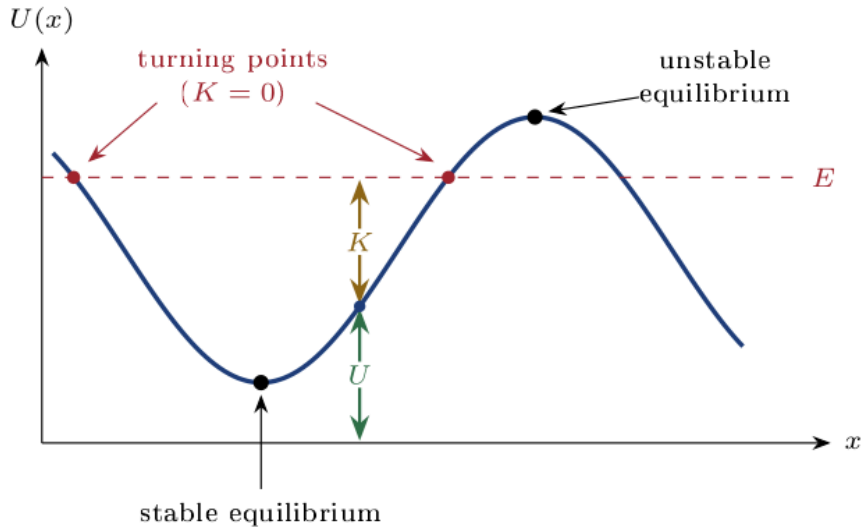


Figure 8.2. An energy diagram. The curve is the potential energy $U(x)$; the dashed line is the particle's total energy E, which never changes. At any position, the vertical gap between the line and the curve is the kinetic energy $K = E - U$. The particle is trapped between the two *turning points*, where the gap closes to nothing. Because the force is the negative slope, it always points downhill on the curve: back toward the valley (stable equilibrium) and away from the peak (unstable equilibrium).

•Figure description

An energy diagram. A curve shows the potential energy U of x, and a horizontal dashed line marks the constant total energy E. The vertical gap between the line and the curve is the kinetic energy. A valley in the curve is marked stable equilibrium and a peak is marked unstable equilibrium, with turning points where the curve meets the line.

Three features of such a diagram do most of the work.

Turning points. Where the E line meets the curve, $K = 0$ and the particle is momentarily at rest. It cannot go further: beyond that point U would exceed E, which would require negative kinetic energy, and there is no such thing. Instead the force turns it around. A particle whose energy line cuts the curve on both sides of a valley is **bound**: it oscillates back and forth between the turning points forever, fastest at the bottom, where U is least and the gap is widest.

Equilibrium. Wherever the curve is flat, $\frac{\mathrm{d}U}{\mathrm{d}x} = 0$, so the force vanishes and a particle placed there at rest stays there. These are the equilibrium positions.

Stability. Not all equilibria are alike, and the shape of the curve tells you which is which. At a *minimum* of U (the bottom of a valley) nudging the particle either way puts it on a slope whose downhill direction points back where it came from. The force restores it, and the equilibrium is **stable**. At a *maximum* of U (the top of a hill) the downhill direction points further away, so any nudge is amplified and the particle runs off. That equilibrium is **unstable**. In calculus terms, $\frac{\mathrm{d}^2U}{\mathrm{d}x^2} > 0$ means stable and $\frac{\mathrm{d}^2U}{\mathrm{d}x^2} < 0$ means unstable. The mental image is a marble in a bowl versus a marble balanced on an upturned bowl. A spring's $U_s = \frac{1}{2}kx^2$ is a parabola with its minimum at $x = 0$, so the relaxed position of a spring is a stable equilibrium, which is why a mass on a spring oscillates about it rather than flying away.

Example 8.15: Describing motion from an energy diagram

A particle is released from rest at the left turning point in Figure 8.2. Describe its subsequent motion without doing any calculation. What would change if its total energy were raised above the peak?

Solution

Released from rest, the particle has $K = 0$ and sits on the curve. The slope there is negative (the curve falls to the right), so by equation (8.8) the force is positive, to the right, and the particle accelerates into the valley, picking up exactly as much kinetic energy as it loses potential.

It is moving fastest at the bottom of the valley, where the gap $E - U$ is widest. Past the bottom the curve rises again, the force reverses, and the particle slows, coming momentarily to rest at the right turning point. There the force pushes it back to the left, and the whole cycle repeats. The particle is trapped, oscillating between the two turning points indefinitely. It cannot reach the peak: doing so would demand more potential energy than it possesses in total.

Raise E above the peak and the picture changes completely. The energy line no longer intersects the curve on the right, so there is no right-hand turning point: the particle climbs over the hill, still slowing near the top, then speeds away down the far side and never returns. It has become **unbound**. The minimum energy needed for this escape is set by the height of the peak, not by the shape of the valley.

Checkpoint 8.6

On an energy diagram, a particle sits exactly at a maximum of $U(x)$, at rest. Strictly speaking it is in equilibrium: the net force on it is zero. Why, then, would you never expect to find a real particle there? And where on the same diagram would you expect to find one?

Answer

Answers to Checkpoints

Checkpoint 8.5. Both marbles arrive at the finish with the same speed: with no friction, each converts the same drop in potential energy, mgh, into kinetic energy, and the shape of the path does not matter. The marble on the dipping track arrives *first*. Its extra drop into the dip makes it go faster than the other marble for most of the trip, which more than makes up for the dip's slightly longer path. The questions are different because energy conservation relates speeds at two *positions*, while arrival time depends on the speed *all along the way*: something the energy equation says nothing about.

Checkpoint 8.6. A maximum of $U(x)$ is an *unstable* equilibrium: the slope is zero there, but on either side the force points away from the peak. The slightest disturbance (a vibration, a molecule's bump) pushes the particle off, and it then accelerates downhill away from the peak. Real particles are found near *minima* of $U(x)$, the stable equilibria, where any small displacement produces a force pushing the particle back. There they sit at rest or oscillate about the bottom of the valley.

This page titled 8.5: Conservation of Mechanical Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.S: Conservation of Energy (Summary)

Summary

- Work done on a system from outside changes the system's energy: $W_{\text{ext}} = \Delta K + \Delta U$ (plus thermal energy if friction acts within the system). Work that changes the system's *motion* goes into kinetic energy; work that changes its *arrangement* goes into potential energy.
- **Kinetic energy** $K = \frac{1}{2}mv^2$ is the scalar, never-negative energy of motion; it grows as the square of speed.
- The **work–kinetic energy theorem**: the net work on an object equals the change in its kinetic energy, $W_{\text{net}} = \Delta K = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$. It is always true, provided W_{net} includes *every* force on the free-body diagram.
- **Potential energy** is stored work, defined so that the work of the internal force is $W_{\text{int}} = -\Delta U$: gravitational $U_g = mgy$ and elastic $U_s = \frac{1}{2}kx^2$. Only changes in U matter, so the reference level is arbitrary and U_g may be negative. Potential energy belongs to a system, not to a single object.
- **Conservative forces** (gravity, the spring force) store energy reversibly; the work they do is path-independent and zero around a closed loop, so a potential energy can be defined for them, with $W_{\text{cons}} = -\Delta U$ and $F_x = -\mathrm{d}U/\mathrm{d}x$. Use either the work term or the U term for a given force, never both.
- **Nonconservative forces** (friction, drag) dissipate mechanical energy, usually as thermal energy; the work they do depends on the path. For kinetic friction, $W_{\text{nc}} = -F_{f,k}d$, where d is the *path length*, not the displacement.
- When only conservative forces do work, **mechanical energy is conserved**: $K_i + U_i = K_f + U_f$. This links the start and end of a process without reference to the path or the time taken; masses often cancel. With nonconservative forces, $\Delta E_{\text{mech}} = W_{\text{nc}}$, giving $K_i + U_i + W_{\text{nc}} = K_f + U_f$. Counting the thermal energy produced, **total energy is always conserved**.
- On an **energy diagram**, $K = E - U$ is the gap between the total-energy line and the curve. The gap closes at **turning points**; the curve is flat at equilibria, which are **stable** at minima of U and **unstable** at maxima.

Key Equations

Energy of a system

$$W_{\text{ext}} = \Delta E_{\text{sys}} = \Delta K + \Delta U \tag{8.S.1}$$

Kinetic energy

$$K = \tfrac{1}{2}mv^2 \tag{8.S.2}$$

Work–kinetic energy theorem

$$W_{\text{net}} = \Delta K = \tfrac{1}{2}mv_f^2 - \tfrac{1}{2}mv_i^2 \tag{8.S.3}$$

Gravitational potential energy

$$U_g = mgy \tag{8.S.4}$$

Elastic potential energy

$$U_s = \tfrac{1}{2}kx^2 \tag{8.S.5}$$

Work done by a conservative force

$$W_{\text{cons}} = -\Delta U = U_i - U_f \tag{8.S.6}$$

Force from a potential-energy function

$$F_x = -\frac{\mathrm{d}U}{\mathrm{d}x} \tag{8.S.7}$$

Change in mechanical energy

$$\Delta E_{\mathrm{mech}} = W_{\mathrm{nc}} \tag{8.S.8}$$

Conservation of mechanical energy

$$K_i + U_i = K_f + U_f \qquad \text{(conservative forces only)} \tag{8.S.9}$$

Energy equation with nonconservative forces

$$K_i + U_i + W_{\mathrm{nc}} = K_f + U_f \tag{8.S.10}$$

Work done by kinetic friction

$$W_{\mathrm{nc}} = -F_{f,k}d, \qquad F_{f,k} = \mu_k F_N \tag{8.S.11}$$

This page titled 8.S: Conservation of Energy (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

8.E: Conservation of Energy (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 8.1 — Where Does the Work Go?

8.1 ★ *[Conceptual.]* In each case a hand does positive work. Say whether the work goes mainly into kinetic energy, potential energy, or thermal energy, and name the system you have in mind.

a. A hockey puck is pushed from rest across frictionless ice.
b. A box is lifted onto a shelf at constant speed.
c. A spring is slowly compressed.
d. A crate is dragged across a rough floor at constant speed.

8.2 ★ *[Conceptual.]* A book is lifted at constant speed from the floor to a table.

a. Taking the book alone as the system, what is the net work done on it, and does its energy change?
b. Taking the book and the Earth as the system, what external work is done, and what kind of energy does the system gain?
c. Explain why the two answers do not contradict each other.

Section 8.2 — Kinetic Energy and the Work–Kinetic Energy Theorem

8.3 ★ Find the kinetic energy of (a) a $0.145\ \text{kg}$ baseball thrown at $40\ \text{m/s}$, and (b) a $1500\ \text{kg}$ car moving at $30\ \text{m/s}$.

8.4 ★★ *[Conceptual.]* A truck and a motorcycle have the same kinetic energy.

a. Which is moving faster, and why?
b. If instead they have the same speed, which has more kinetic energy?

8.5 ★★ An electron of mass $9.11 \times 10^{-31}\ \text{kg}$ moves at $2.0 \times 10^{6}\ \text{m/s}$. Find its kinetic energy in joules, and comment on the size of the number.

8.6 ★★ A $0.45\ \text{kg}$ ball is thrown straight up at $18\ \text{m/s}$.

a. Using the work–kinetic energy theorem, find the height at which its speed has dropped to $9.0\ \text{m/s}$. (Neglect air resistance.)
b. What is the net work done on the ball between launch and that height?

8.7 ★★ *[Real-world: transportation.]* A $1100\ \text{kg}$ car traveling at $25\ \text{m/s}$ brakes to a stop.

a. How much work do the brakes (and friction) do on the car?
b. If the same car is traveling at twice the speed, how does the stopping distance compare, assuming the same braking force? Explain using the work–kinetic energy theorem.

8.8 ★★ Return to the force–position graph of Problem 7.18. The object has mass $2.0\ \text{kg}$, starts from rest at $x = 0$, and no other force does work on it. How fast is it moving at $x = 2\ \text{m}$, and at $x = 5\ \text{m}$?

8.9 ★★ A $30\ \text{kg}$ sled starts from rest on level snow ($\mu_k = 0.20$) and is pulled by a rope with a tension of $80\ \text{N}$ directed 25° above the horizontal.

a. Draw the free-body diagram and find the normal force and the friction force.
b. Find the work done by each force over the first $10\ \text{m}$, and the net work.
c. Use the work–kinetic energy theorem to find the sled's speed after $10\ \text{m}$.

8.10 ★★ *[Real-world: forensics.]* After a collision, an accident investigator measures skid marks $40\ \text{m}$ long left by a car that locked its brakes on a level road. The coefficient of kinetic friction for rubber on that pavement is 0.70.

a. Use the work–kinetic energy theorem to estimate the car's speed when the skid began. Does your answer depend on the car's mass?
b. On wet pavement ($\mu_k = 0.65$), how long would the skid be for a car traveling at $28\ \text{m/s}$?

8.11 ★★ Return to the crate of Problem 7.14, released from rest on a 25° ramp. How fast is it moving when its energy has increased by $100\ \text{J}$? Check your answer with Newton's second law and the kinematics of Chapter 2.

8.12 ★★★ A $2.0\ \text{kg}$ block is pushed up a frictionless 20° incline by a force of $15\ \text{N}$ directed up the slope. Starting from rest, use the work–kinetic energy theorem to find its speed after it has traveled $1.5\ \text{m}$ along the incline. *(Account for the work done by both the applied force and gravity.)*

Section 8.3 — Potential Energy

8.13 ★ A $2.0\ \text{kg}$ textbook is lifted from the floor to a shelf $1.8\ \text{m}$ high.

a. What is its gravitational potential energy relative to the floor?
b. What would it be relative to a desk $0.75\ \text{m}$ high?
c. Explain why both answers are acceptable.

8.14 ★ *[Conceptual.]* Two identical balls are at the same height, one resting on a shelf and one moving upward past the shelf at $4\ \text{m/s}$.

a. Do they have the same gravitational potential energy? Explain.
b. Do they have the same total mechanical energy? Explain.

8.15 ★★ A spring with force constant $320\ \text{N/m}$ is compressed $0.12\ \text{m}$.

a. How much elastic potential energy is stored?
b. If a $0.50\ \text{kg}$ ball is placed against the spring and released, how fast will it be moving when the spring reaches its relaxed length (assume a frictionless, horizontal surface)? *(Use the work–kinetic energy theorem: the work the spring does on the ball is the energy that was stored in it.)*

8.16 ★★ Return to the mass of Problem 7.20, lowered slowly onto a spring until it hangs at rest.

a. By how much does the gravitational potential energy change?
b. How much elastic potential energy is stored in the spring at the end?
c. The two answers do not add to zero. Where did the difference go?

8.17 ★ *[Real-world: hiking.]* A $65\ \text{kg}$ hiker descends from the rim of a canyon to the canyon floor, $120\ \text{m}$ below. Taking the rim as the reference level, what is the hiker's gravitational potential energy on the canyon floor? Is a negative answer a problem? Explain.

Section 8.4 — Conservative and Nonconservative Forces

8.18 ★ *[Conceptual.]* Classify each force as conservative or nonconservative, and give a one-sentence reason: (a) gravity, (b) the force of a spring, (c) sliding (kinetic) friction, (d) air resistance.

8.19 ★ *[Conceptual.]* You carry a suitcase from the ground floor to the third floor of a building.

a. Does the work done by gravity depend on whether you take the stairs or the elevator? Explain.
b. Does the work *you* do against gravity depend on the route? What does this tell you about gravity being conservative?

8.20 ★★ A $2.0\ \text{kg}$ block is pushed around a closed rectangular loop on a level floor and returns to its starting point.

a. What is the total work done by gravity over the loop?
b. If the coefficient of kinetic friction is 0.30 and the loop's perimeter is $6.0\ \text{m}$, how much mechanical energy is dissipated by friction?

8.21 ★★ *[Conceptual.]* A ball is dropped and bounces several times, returning to a lower height after each bounce until it stops.

a. Is mechanical energy conserved during this process? Explain.
b. Is total energy conserved? Where does the "lost" mechanical energy go?

8.22 ★★ A $4.0\ \text{kg}$ object slides $3.0\ \text{m}$ down a rough incline, and friction removes $22\ \text{J}$ of mechanical energy. Find the friction force, assuming it is constant.

8.23 ★★ The potential energy of a particle moving along the x-axis is $U(x) = 5.0x^2 - 2.0x^3$, with U in joules and x in meters.

a. Find the force $F_x(x)$ on the particle.
b. Find the force at $x = 1.0\ \text{m}$ and at $x = 2.0\ \text{m}$, and give the direction of each.
c. At what positions is the force zero?

8.24 ★★★ A block slides down a curved, frictionless ramp from a height h and then continues onto a rough horizontal surface where it eventually stops. Explain, using equations (8.10) and (8.9), why the stopping distance on the flat depends on h and the coefficient of friction but *not* on the shape of the ramp.

Section 8.5 — Conservation of Mechanical Energy

8.25 ★ A $0.25\ \text{kg}$ ball is dropped from rest at a height of $12\ \text{m}$.

a. Use conservation of energy to find its speed just before it hits the ground.
b. Does the answer depend on the ball's mass? Explain.

8.26 ★ A $60\ \text{kg}$ skateboarder starts from rest at the top of a frictionless ramp $4.0\ \text{m}$ high. How fast is she moving at the bottom?

8.27 ★★ A pendulum bob is released from rest with its string horizontal. The string is $1.2\ \text{m}$ long.

a. How fast is the bob moving at the lowest point?
b. How fast is it moving when the string makes a 45° angle with the vertical?

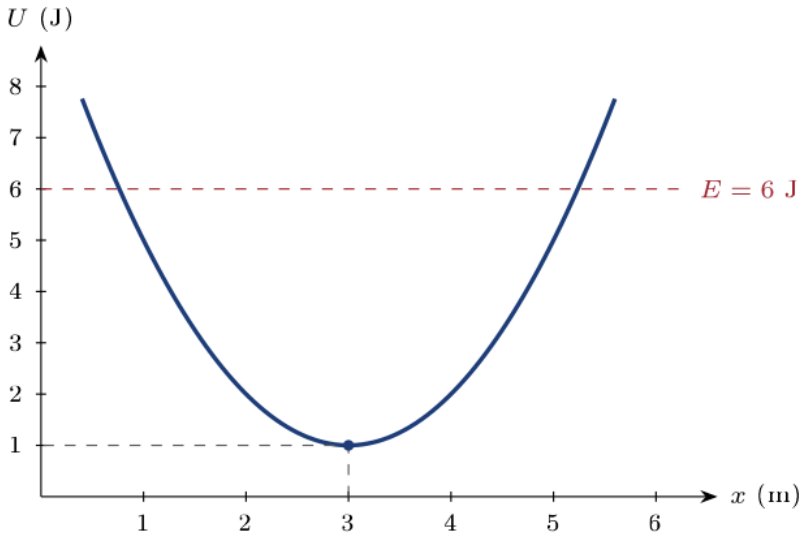
8.28 ★★ *[Real-world: sports.]* A $55\ \text{kg}$ pole-vaulter converts the kinetic energy of her $9.0\ \text{m/s}$ sprint entirely into height.

a. Ignoring the pole and her own push, how high can her center of mass rise?
b. Real vaulters clear bars higher than this estimate. Suggest one reason why.

8.29 ★★ A spring of force constant $800\ \text{N/m}$ is compressed $0.20\ \text{m}$ and used to launch a $0.50\ \text{kg}$ ball horizontally off the edge of a frictionless table.

a. Find the ball's speed as it leaves the spring.
b. If the table is $1.0\ \text{m}$ high, how far from the base of the table does the ball land? *(Combine energy methods with projectile motion from Chapter 4.)*

8.30 ★★ *[Graph-reading.]* The graph shows the potential energy $U(x)$ of a $1.0\ \text{kg}$ particle moving along the x-axis. The particle has a total mechanical energy of $6.0\ \text{J}$.



•Figure description
A graph of potential energy in joules versus position in meters for a one kilogram particle, with a horizontal dashed line marking the total mechanical energy of six joules.

a. At a point where $U = 2.0$ J, what is the particle's kinetic energy and speed?
b. What is the particle's speed where U is a minimum (read the minimum value from the graph)?
c. Explain what happens at a point where $U = 6.0$ J. Between what two values of x is the particle confined?
d. Identify the equilibrium position, and state whether it is stable or unstable. Give your reason in terms of the shape of the curve.

8.31 ★ *[Conceptual.]* A roller-coaster car coasts (engine off) from the top of the first hill. Explain why no subsequent hill on a frictionless track can be taller than the first.

8.32 ★★★ A 2.0 kg block slides down a frictionless track from rest at height 1.5 m, then compresses a spring ($k = 600$ N/m) at the bottom.

a. Find the block's speed just before it reaches the spring.
b. Find the maximum compression of the spring.

8.33 ★★ A 1200 kg car coasts in neutral down a 6.0 % grade (dropping 6.0 m per 100 m of road) and reaches a steady speed, with all the gravitational energy going to friction and drag. Over 500 m of road, how much mechanical energy is dissipated?

8.34 ★★ *[Real-world: engineering.]* Water flows over a dam, dropping 45 m at a rate of 1.5×10^5 kg/s.

a. How much gravitational potential energy is released per second?
b. If a turbine converts 85 % of this into electrical power, what is the power output in megawatts?

8.35 ★★★ A 0.50 kg block is pressed against a spring ($k = 400$ N/m), compressing it 0.15 m, then released. It travels 0.80 m along a rough horizontal surface ($\mu_k = 0.25$) before reaching the bottom of a frictionless ramp.

a. How much energy does the spring store?
b. How much mechanical energy is lost to friction before the ramp?
c. How high up the ramp does the block rise?

8.36 ★★★ *[Conceptual.]* A skier of mass m descends a frictionless slope of height h, and a second skier of mass $2m$ descends the same slope.

a. Compare their speeds at the bottom.
b. Compare their kinetic energies at the bottom.
c. Explain why one comparison depends on mass and the other does not.

8.37 ★ A $0.060\ \text{kg}$ tennis ball is dropped from a height of $2.0\ \text{m}$ and rebounds to a height of $1.4\ \text{m}$.

a. How much mechanical energy was lost in the bounce?
b. What fraction of the ball's original mechanical energy is this?

Challenge Problems

8.38 ★★★ Return to the force $F(x) = (12 - 3x^2)\ \text{N}$ of Problem 7.32, acting on a $1.5\ \text{kg}$ object that starts from rest at the origin on a frictionless surface.

a. Find the object's speed at $x = 1.0\ \text{m}$.
b. Find its maximum speed, and where it occurs.
c. Find the position at which it momentarily comes to rest again.
d. Is $F(x)$ a conservative force? If so, find a potential-energy function $U(x)$ for it, taking $U(0) = 0$, and use it to check your answer to (c).

8.39 ★★★ A spring of force constant $500\ \text{N/m}$ is compressed $0.20\ \text{m}$ and used to launch a $3.0\ \text{kg}$ block along a horizontal surface. The short launch region is frictionless, but beyond it the surface has $\mu_k = 0.25$.

a. How much elastic potential energy is stored in the compressed spring?
b. How fast is the block moving as it leaves the spring?
c. How far along the rough surface does the block slide before stopping?
d. If instead the rough surface were only $0.80\ \text{m}$ long (and frictionless thereafter), what would the block's final speed be?

8.40 ★★★ *[Real-world: transportation.]* A $1200\ \text{kg}$ car accelerates from rest with its engine delivering a constant useful power of $90\ \text{kW}$. Neglect air resistance and friction.

a. Using the work–kinetic energy theorem, show that the car's speed grows as $v(t) = \sqrt{2Pt/m}$.
b. How long does it take to reach $30\ \text{m/s}$?
c. How far does the car travel in that time?
d. Explain why the car's acceleration decreases as it speeds up, even though the power is constant.

8.41 ★★★ *[Real-world: engineering.]* An engineer tests a roller-coaster design on a scale model: a $0.20\ \text{kg}$ car runs on a smooth track and passes through a vertical circular loop of radius $R = 0.40\ \text{m}$. The car is released from rest at a height h above the bottom of the loop. *(You will need uniform circular motion from Chapter 6.)*

a. What is the minimum speed the car must have at the top of the loop to stay in contact with the track?
b. Find the minimum release height h_{min} (in terms of R, then numerically) needed to complete the loop.
c. If the car is instead released from $h = 1.5\ \text{m}$, find its speed at the top of the loop and the normal force the track exerts on it there.

8.42 ★★★ A $1.5\ \text{kg}$ block is launched up a rough incline ($\theta = 25^\circ$) with an initial speed of $6.0\ \text{m/s}$. The coefficients of friction are $\mu_k = 0.30$ and $\mu_s = 0.40$.

a. How far up the incline does the block travel before momentarily stopping?
b. Will the block then slide back down? Justify with the appropriate friction condition.
c. If it slides back, how fast is it moving when it returns to its starting point?

8.43 ★★★ A $0.50\ \text{kg}$ particle moves along the x-axis with potential energy $U(x) = 2x^3 - 9x^2 + 12x$, where U is in joules and x in meters.

a. Find the force $F(x)$ acting on the particle.
b. Locate the equilibrium positions.
c. Classify each equilibrium as stable or unstable, with a reason.
d. Sketch $U(x)$ for $0 \le x \le 3$ and mark both equilibria on your sketch.
e. If the particle rests at the stable equilibrium, what minimum speed must it be given to escape the potential well, and in which direction must it be sent? Explain why the other direction will not do.

This page titled 8.E: Conservation of Energy (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9: LINEAR MOMENTUM AND CENTER OF MASS

When two cars collide, when a rifle kicks back against a shoulder, when a rocket climbs by throwing gas out its nozzle, the forces involved are violent, complicated, and usually unknown in detail. Yet the outcomes obey a law so simple and so powerful that we can predict them without knowing any of those forces. That law is the conservation of **linear momentum**, and it is the subject of this chapter.

Momentum is, loosely, “inertia in motion”: a measure of how much motion an object carries and how hard it is to stop. Newton himself wrote his second law not as $\mathbf{F} = m\mathbf{a}$ but in terms of momentum, and that original form turns out to be the more general and more useful one. From it we will build the idea of **impulse** (the momentum delivered by a force over time), the conservation principle that governs all interactions, and the tools to analyze collisions in one and two dimensions. Finally we will meet the **center of mass**, the single point that moves as though all the external forces acted on all the mass at once: the key that lets us treat a whole system of particles as if it were one.

Learning Objectives

- Define linear momentum and express Newton’s second law in terms of it.
- Relate the impulse of a force to the change in an object’s momentum.
- Apply conservation of linear momentum to interactions where external forces are negligible.
- Distinguish elastic, inelastic, and perfectly inelastic collisions, and solve them in one and two dimensions.
- Locate the center of mass of a system of particles and describe how it moves.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

9.1: Linear Momentum
9.2: Impulse
9.3: Conservation of Linear Momentum
9.4: Collisions
9.5: Center of Mass
9.S: Linear Momentum and Center of Mass (Summary)
9.E: Linear Momentum and Center of Mass (Problems)

This page titled 9: Linear Momentum and Center of Mass is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.1: Linear Momentum

The **linear momentum** of an object is the product of its mass and its velocity:

Key Equation: Linear momentum

$$\mathbf{p} = m\mathbf{v}. \tag{9.1}$$

Momentum is a vector, pointing in the same direction as the velocity, with SI units of $\text{kg} \cdot \text{m/s}$. A bowling ball rolling at walking pace ($7.0\ \text{kg}$ at $1.4\ \text{m/s}$) and a baseball leaving a pitcher's hand ($0.145\ \text{kg}$ at $40\ \text{m/s}$) carry momenta of $9.8\ \text{kg} \cdot \text{m/s}$ and $5.8\ \text{kg} \cdot \text{m/s}$: the same order of magnitude, despite a fifty-fold difference in mass. What matters is the product mv, not either factor alone.

It is often useful to write kinetic energy in terms of momentum rather than speed. Since $p = mv$,

$$K = \tfrac{1}{2}mv^2 = \frac{p^2}{2m}. \tag{9.2}$$

Written this way, K makes plain something that momentum and kinetic energy do *not* share: at a given momentum, a light object carries far more kinetic energy than a heavy one. We will put this to work in §9.3.

Momentum is more than a convenient combination of symbols: it is the quantity at the heart of Newton's second law. In its original and most general form, the net force on an object equals the rate of change of its momentum:

Key Equation: Newton's second law (momentum form)

$$\mathbf{F}_{\text{net}} = \frac{d\mathbf{p}}{dt}. \tag{9.3}$$

When the mass is constant this reduces to the familiar form, since $\frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} = m\frac{d\mathbf{v}}{dt} = m\mathbf{a}$. But equation (9.3) is the more fundamental statement, and it points directly to the conservation law we are after: if the net external force is zero, then $d\mathbf{p}/dt = 0$ and the momentum cannot change at all.

Note: A caution about variable-mass systems

Equation (9.3) applies to a *fixed collection of matter*, not to a container whose contents are streaming away. A rocket is the classic trap: setting $\mathbf{F}_{\text{net}} = d(m\mathbf{v})/dt$ for the rocket alone, with m shrinking as fuel burns, gives the wrong answer, because the exhaust carries momentum out of the accounting. The remedy is to choose the system so that no matter crosses its boundary (the rocket *plus* the parcel of gas it is about to expel) and then apply momentum conservation to that fixed collection. You will carry out exactly this argument in the rocket problem at the end of the chapter.

This page titled 9.1: Linear Momentum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.2: Impulse

A momentum changes only when a force acts, and the change depends on both how strong the force is and how long it acts. Rearranging equation (9.3) as $d\mathbf{p} = \mathbf{F}\ dt$ and adding up the contributions over the time of the interaction gives the **impulse** $\mathbf{J}$, which equals the change in momentum:

Key Equation: Impulse–momentum theorem

$$\mathbf{J} = \int \mathbf{F}\, dt = \mathbf{F}_{\text{avg}}\, \Delta t = \Delta \mathbf{p}. \tag{9.4}$$

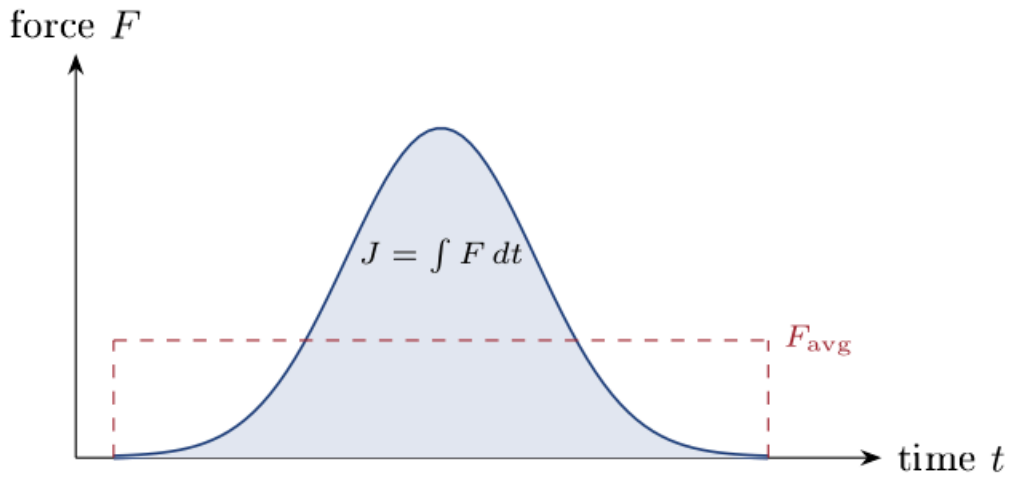


Figure 9.1. The impulse delivered by a force is the area under its force–time graph. The average force F_{avg} is the height of the rectangle (over the same time interval) with the same area.

•Figure description

A graph of force versus time for a collision. The area under the curved force pulse is the impulse. A dashed rectangle of the same area is drawn, its height marking the average force.

Impulse has units of $\text{N} \cdot \text{s}$, which are the same as the units of momentum: $\text{N} \cdot \text{s} = (\text{kg} \cdot \text{m/s}^2)(\text{s}) = \text{kg} \cdot \text{m/s}$. That the two agree is a good sign, since equation (9.4) sets them equal. Graphically, the impulse is the area under the force-versus-time curve (Figure 9.1); the average force is the constant force that would enclose the same area over the same time.

The impulse–momentum theorem explains a great deal of everyday safety design. For a fixed change in momentum (catching a ball, stopping in a crash) the impulse is fixed, so a longer contact time means a smaller force. Airbags, crumple zones, padded dashboards, and bending your knees when you land all work the same way: they stretch out Δt to shrink the force.

Checkpoint 9.1

A catcher must absorb the same momentum change from a given fastball no matter how the catch is made. Why, then, does drawing the glove backward during the catch reduce the sting? Which quantity in the impulse–momentum theorem is the catcher actually controlling?

Answer

Example 9.1: Crash test

In a crash test, a $1500\ \text{kg}$ car strikes a wall. Its velocity is $-15.0\ \text{m/s}$ just before impact and $+2.60\ \text{m/s}$ (it rebounds) just after. If the collision lasts $0.150\ \text{s}$, find the impulse on the car and the average net force exerted on it.

Solution

Work along the direction of motion, with the direction of rebound (away from the wall) positive, so the incoming velocity is negative.

Impulse. The impulse equals the change in momentum, final minus initial:

$$\begin{aligned} J = \Delta p = m(v_f - v_i) &= (1500\,\text{kg})\big(2.60 - (-15.0)\big)\,\text{m/s} \\ &= (1500)(17.6)\,\text{kg}\cdot\text{m/s} = \boxed{2.64 \times 10^4\,\text{N}\cdot\text{s}}. \end{aligned} \tag{9.2.1}$$

Subtracting the negative initial velocity *adds* its size: the car must first lose $15.0\ \text{m/s}$ of motion toward the wall and then gain $2.60\ \text{m/s}$ of motion away from it, a total change of $17.6\ \text{m/s}$. A rebound always requires a bigger impulse than simply stopping. (The units agree because $\text{kg}\cdot\text{m/s} = (\ \text{kg}\cdot\text{m/s}^2)\ \ \text{s} =\ \ \text{N}\cdot\text{s}$.)

The impulse is positive: directed away from the wall.

Average force. Solve $J = F_{\text{avg}}\ \Delta t$ for the force by dividing both sides by Δt:

$$F_{\text{avg}} = \frac{J}{\Delta t} = \frac{2.64 \times 10^4\,\text{N}\cdot\text{s}}{0.150\,\text{s}} \approx \boxed{1.76 \times 10^5\,\text{N}}. \tag{9.2.2}$$

That is about $(1.76 \times 10^5)/(1.47 \times 10^4) \approx 12$ times the car's own weight ($mg = 1.47 \times 10^4\ \text{N}$), sustained for just over a tenth of a second.

Answers to Checkpoints

Checkpoint 9.1. Drawing the glove back stretches out the time Δt over which the ball is stopped. The impulse $J = \Delta p$ is fixed by the ball's mass and speed, but since $J = F_{\text{avg}}\ \Delta t$, a longer stopping time means a proportionally smaller average force on the hand. The catcher cannot change Δp; what the catcher controls is Δt, and through it F_{avg}.

This page titled 9.2: Impulse is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.3: Conservation of Linear Momentum

Equation (9.3) carries an immediate and profound consequence. When a system contains more than one object, write its total momentum as the vector sum of the individual momenta,

$$\mathbf{P} = \sum_i \mathbf{p}_i = \sum_i m_i \mathbf{v}_i, \tag{9.5}$$

using a capital P for the system total to keep it distinct from the momentum $\mathbf{p}$ of any single object. If the net *external* force on the system is zero, $\mathbf{P}$ cannot change:

Key Equation: Conservation of linear momentum

$$\text{if } \mathbf{F}_{\text{ext,net}} = 0, \qquad \mathbf{P}_i = \mathbf{P}_f \quad (\text{total momentum constant}). \tag{9.6}$$

The internal forces between parts of a system (however large) always come in Newton's-third-law pairs that cancel in the total, so they cannot change $\mathbf{P}$. Only an outside force can. This is what makes momentum so useful for collisions and explosions: during the brief, intense interaction the internal forces dwarf gravity and friction, so to good approximation the total momentum just before equals the total momentum just after, even though we never need to know the messy details of the force between the objects.

A rifle and bullet make the cleanest illustration. Before firing, the total momentum is zero. The expanding gas pushes the bullet forward and, with equal and opposite total force, pushes back on the rifle; afterward the forward momentum of the bullet is exactly balanced by the backward momentum of the recoiling rifle, keeping the total at zero.

Example 9.2: Rifle recoil

A $3.6\ \text{kg}$ rifle fires a $9.5\ \text{g}$ bullet at $720\ \text{m/s}$. Find the recoil velocity of the rifle, and compare the kinetic energy carried off by the bullet with that of the recoiling rifle.

Solution

Choosing the system. Take the rifle and bullet *together* as the system, with the bullet's direction of travel as positive. The gas pressure that drives them apart is internal to this system; gravity and the shooter's grip are external, but they act for only the few milliseconds of firing, so their impulse is negligible compared with the momentum exchanged. Total momentum is therefore conserved, and it starts at zero, since nothing is moving before the shot.

Recoil velocity. Convert the bullet's mass to kilograms first: $9.5\ \text{g} = 0.0095\ \text{kg}$. Setting the total momentum after firing equal to zero, subtract $m_b v_b$ from both sides and divide by m_R:

$$0 = m_b v_b + m_R v_R \implies m_R v_R = -m_b v_b \implies v_R = -\frac{m_b v_b}{m_R}, \tag{9.3.1}$$

so

$$v_R = -\frac{(0.0095\,\text{kg})(720\,\text{m/s})}{3.6\,\text{kg}} = \boxed{-1.9\,\text{m/s}}, \tag{9.3.2}$$

the minus sign meaning the rifle moves backward. The two momenta have equal magnitude, $(0.0095)(720) = 6.8\ \text{kg}\cdot\text{m/s}$ apiece, and opposite signs, so they add to zero.

Kinetic energies. These are nothing like equal. Using $K = p^2/2m$ from equation (9.2), and noting that the bullet and rifle carry the same $|p|$, write the ratio. Dividing by a fraction is multiplying by its reciprocal, and then p^2 and the 2 cancel:

$$\frac{K_b}{K_R} = \frac{p^2/2m_b}{p^2/2m_R} = \frac{p^2}{2m_b} \times \frac{2m_R}{p^2} = \frac{m_R}{m_b} = \frac{3.6}{0.0095} \approx \boxed{380}. \tag{9.3.3}$$

Numerically, $K_b = \frac{1}{2}(0.0095)(720)^2 = 2.46 \times 10^3$ J while $K_R = \frac{1}{2}(3.6)(1.9)^2 = 6.50$ J, and indeed $2460/6.50 \approx 380$. Momentum is split evenly between the two objects; energy is not, and it lands overwhelmingly on the lighter one. That single fact is why the recoil only bruises a shoulder. It is also a warning: momentum and kinetic energy are built from the same m and v, but they answer very different questions, and a system can conserve one while losing most of the other.

Checkpoint 9.2

A ball is dropped from rest and falls toward the ground. Is the ball's momentum conserved as it falls? Is the momentum of the ball–Earth system conserved? What does your answer imply about the Earth?

Answer

Answers to Checkpoints

Checkpoint 9.2. The ball's momentum is *not* conserved: gravity is an external force on the ball alone, and its downward momentum grows steadily. The ball–Earth system's momentum *is* conserved, because the gravitational forces between them are internal to that system (and third-law partners). So as the ball gains downward momentum, the Earth must gain an equal upward momentum. Because the Earth's mass is some 10^{24} times larger, its velocity change is immeasurably small, but it is there.

This page titled 9.3: Conservation of Linear Momentum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.4: Collisions

A **collision** is a brief interaction in which objects exchange momentum. Because external forces are negligible during the impact, total momentum is conserved in *every* collision. What distinguishes one collision from another is what happens to the kinetic energy:

- In an **elastic** collision, kinetic energy is also conserved. Billiard balls and atoms come close to this ideal.
- In an **inelastic** collision, some kinetic energy is lost (to heat, sound, deformation); momentum is still conserved.
- In a **perfectly inelastic** collision, the objects stick together and move with a common velocity afterward. This loses the most kinetic energy consistent with conserving momentum.

Collisions in one dimension

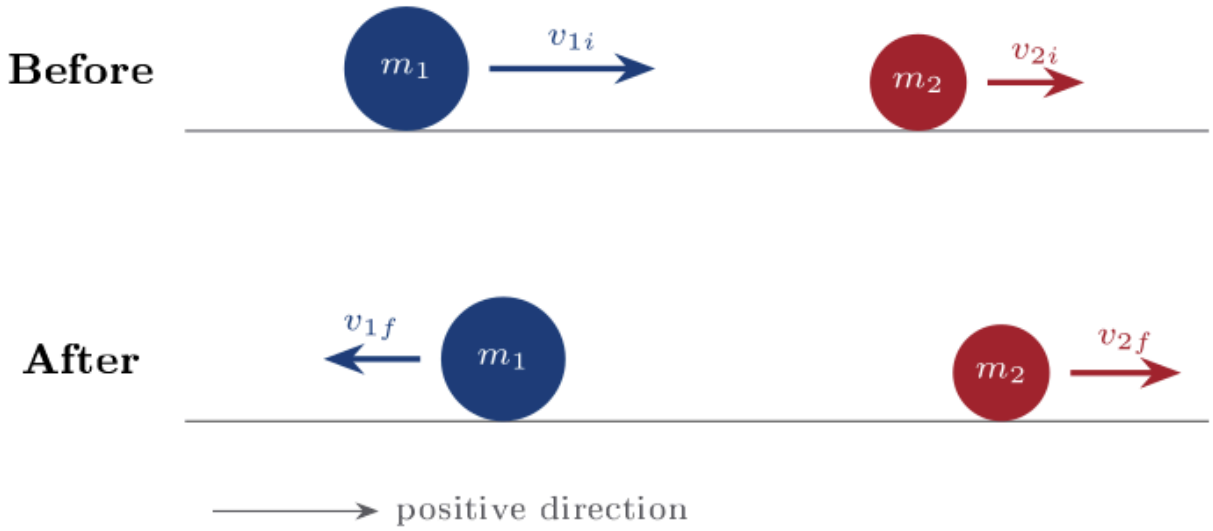


Figure 9.2. The before-and-after picture for a one-dimensional collision. Once a positive direction is chosen, every velocity is a signed number: here v_{1f} is negative, because m_1 rebounds. Conservation of momentum, equation (9.7), relates the top row to the bottom row.

•Figure description

A before-and-after diagram of a one-dimensional collision. In the upper row, masses m sub 1 and m sub 2 approach with labeled initial velocities; in the lower row they move off with their final velocities. A chosen positive direction makes each velocity a signed number.

Before writing any equation, draw the before-and-after picture (Figure 9.2). In one dimension the vector arrows can be dropped, but only if the signs are kept: choose a positive direction once, at the outset, and then write every velocity as a *signed* number, negative for an object moving the other way. Far more wrong answers in this section come from a dropped minus sign than from a wrong equation.

For all collisions, momentum is conserved. Let us consider a collision between two objects. The momentum of both objects combined must be the same both before and after the collision:

$$m_1\mathbf{v}_{1i} + m_2\mathbf{v}_{2i} = m_1\mathbf{v}_{1f} + m_2\mathbf{v}_{2f}. \tag{9.7}$$

In a perfectly inelastic collision, the final velocities of the objects are the same: $\mathbf{v}_{1f} = \mathbf{v}_{2f} = \mathbf{v}_f$. Applying this to (9.7) gives

$$m_1\mathbf{v}_{1i} + m_2\mathbf{v}_{2i} = (m_1 + m_2)\mathbf{v}_f. \tag{9.8}$$

For a one-dimensional *elastic* collision we have two conditions: conservation of momentum and conservation of kinetic energy. This gives two equations to work with, and thus two unknowns can be solved for. Conservation of kinetic energy gives

$$\frac{1}{2}m_1v_{1i}^2+\frac{1}{2}m_2v_{2i}^2=\frac{1}{2}m_1v_{1f}^2+\frac{1}{2}m_2v_{2f}^2. \tag{9.9}$$

Solving this system of two equations can be quite tedious due to the quadratic nature of the kinetic energy. However, it can be shown that (9.7) and (9.9) together imply a much simpler relation:

$$v_{1i}-v_{2i}=-(v_{1f}-v_{2f}). \tag{9.10}$$

In words: in a one-dimensional elastic collision the two objects separate at the same rate at which they approached. Equation (9.10) is linear in the final velocities, so solving it together with (9.7) is straightforward algebra. Doing so gives

Key Equation: Elastic collisions in one dimension

$$v_{1f}=\frac{m_1-m_2}{m_1+m_2}\,v_{1i}+\frac{2m_2}{m_1+m_2}\,v_{2i},\qquad v_{2f}=\frac{2m_1}{m_1+m_2}\,v_{1i}+\frac{m_2-m_1}{m_1+m_2}\,v_{2i}. \tag{9.11}$$

Example 9.3: A rear-end collision

An $1800\ \text{kg}$ car is stopped at a light when a $900\ \text{kg}$ car traveling at $20.0\ \text{m/s}$ strikes it from behind. The two become entangled and move off together. Find their common velocity just after impact.

Solution

The cars stick together, so this is a perfectly inelastic collision: momentum is conserved, and afterward both cars share a single velocity v_f. With the stopped car (m_1) at rest, its initial momentum is zero, and equation (9.8) reduces to

$$m_2v_{2i}=(m_1+m_2)v_f. \tag{9.4.1}$$

Divide both sides by the total mass (m_1+m_2) to isolate v_f:

$$v_f=\frac{m_2v_{2i}}{m_1+m_2}=\frac{(900\,\text{kg})(20.0\,\text{m/s})}{1800\,\text{kg}+900\,\text{kg}}=\frac{18000}{2700}\ \text{m/s}\approx\boxed{6.67\,\text{m/s}}, \tag{9.4.2}$$

in the direction the moving car was traveling. The same momentum is now shared by three times the mass, so the speed is a third of $20.0\ \text{m/s}$.

Energy. Kinetic energy is *not* conserved. Before, only the moving car has any: $K_i=\frac{1}{2}(900)(20.0)^2=1.80\times10^5$ J. After, $K_f=\frac{1}{2}(2700)(6.67)^2=6.00\times10^4$ J. The fraction remaining is $K_f/K_i=6.00/18.0=1/3$, so the collision loses $2/3$ of the kinetic energy to crumpled metal, heat, and sound. In general, for a moving object striking one at rest and sticking, the fraction lost is $m_1/(m_1+m_2)=1800/2700=2/3$. Note that how much is lost depends on the mass ratio: it is not always more than half.

Example 9.4: An elastic collision via a spring

A block $m_1=1.60\ \text{kg}$ moves right at $4.00\ \text{m/s}$ on a frictionless track and collides with a light spring attached to a block $m_2=2.10\ \text{kg}$ moving left at $2.50\ \text{m/s}$. The spring ($k=600\ \text{N/m}$) makes the collision elastic. Find (a) the velocity of each block after they separate, and (b) the maximum compression of the spring during the collision.

Solution

Take rightward as positive: $v_{1i}=+4.00\ \text{m/s}$, $v_{2i}=-2.50\ \text{m/s}$. The minus sign on v_{2i} is essential (block 2 is moving left), and it must be carried into every formula. The spring stores and returns the kinetic energy, so the collision is elastic and equations (9.11) apply, with $m_1+m_2=3.70\ \text{kg}$.

(a) Final velocities. Evaluate each fraction first, then multiply by the velocity it goes with:

$$\frac{m_1-m_2}{m_1+m_2}=\frac{1.60-2.10}{3.70}=\frac{-0.50}{3.70}=-0.135,\qquad\frac{2m_2}{m_1+m_2}=\frac{4.20}{3.70}=1.135, \tag{9.4.3}$$

so

$$v_{1f} = (-0.135)(4.00) + (1.135)(-2.50) = -0.541 - 2.838 \approx \boxed{-3.38\,\text{m/s}}. \tag{9.4.4}$$

Both terms are negative: block 1 is the lighter of the two, so it bounces back off the heavier block, and the incoming block 2 was also moving left. Similarly,

$$\frac{2m_1}{m_1+m_2} = \frac{3.20}{3.70} = 0.865, \qquad \frac{m_2-m_1}{m_1+m_2} = \frac{0.50}{3.70} = 0.135, \tag{9.4.5}$$

so

$$v_{2f} = (0.865)(4.00) + (0.135)(-2.50) = 3.459 - 0.338 \approx \boxed{+3.12\,\text{m/s}}. \tag{9.4.6}$$

Block 1 rebounds to the left and block 2 moves off to the right.

Check: momentum. Before: $(1.60)(4.00) + (2.10)(-2.50) = 6.40 - 5.25 = 1.15\ \text{kg}\cdot\text{m/s}$. After: $(1.60)(-3.38) + (2.10)(3.12) = -5.41 + 6.56 = 1.15\ \text{kg}\cdot\text{m/s}$. ✓

Check: kinetic energy. Before: $\frac{1}{2}(1.60)(4.00)^2 + \frac{1}{2}(2.10)(2.50)^2 = 12.80 + 6.56 = 19.36$ J. After: $\frac{1}{2}(1.60)(3.38)^2 + \frac{1}{2}(2.10)(3.12)^2 = 9.14 + 10.22 = 19.36$ J. ✓ (Squaring removes the minus signs, so energy never depends on direction.)

(b) Maximum compression. For part (b) we finally need k. While the blocks approach each other the spring is being squeezed; once they begin to separate, it is relaxing. The spring is therefore at maximum compression at the instant the blocks stop approaching each other, that is, when they momentarily share a common velocity. At that instant they move together, like a perfectly inelastic collision, so momentum conservation fixes that velocity:

$$v_{\text{common}} = \frac{P}{m_1+m_2} = \frac{1.15\,\text{kg}\cdot\text{m/s}}{3.70\,\text{kg}} = 0.311\,\text{m/s}. \tag{9.4.7}$$

Mechanical energy is still conserved (the spring is ideal), so whatever kinetic energy has gone missing at that instant is stored in the spring:

$$U_s = K_i - K_{\text{common}} = 19.36 - \tfrac{1}{2}(3.70)(0.311)^2 = 19.36 - 0.18 = 19.18\,\text{J}. \tag{9.4.8}$$

Setting $U_s = \frac{1}{2}kx^2$ and solving for x (multiply by 2, divide by k, take the square root):

$$x = \sqrt{\frac{2U_s}{k}} = \sqrt{\frac{2(19.18\,\text{J})}{600\,\text{N/m}}} = \sqrt{0.0639\,\text{m}^2} \approx \boxed{0.253\,\text{m}}. \tag{9.4.9}$$

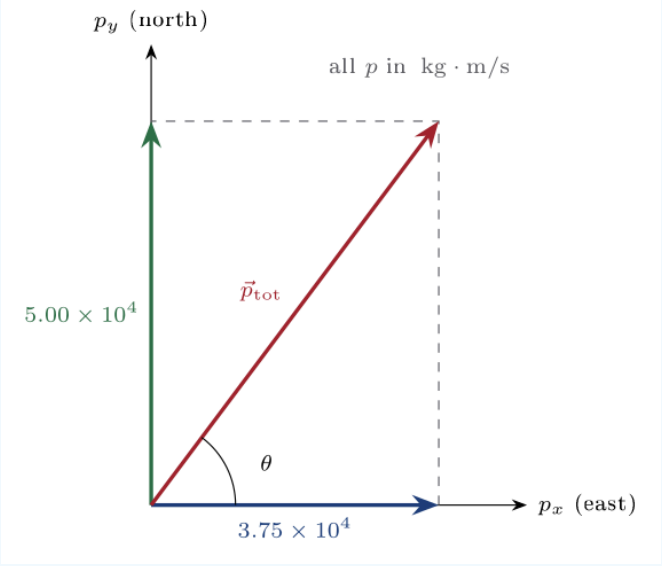
Note the division of labor: momentum and energy conservation together fixed the final velocities in part (a), but finding what happens *during* the collision took one more physical idea: that maximum compression means a common velocity.

Collisions in two dimensions

Collisions need not be head-on. In two dimensions momentum is still conserved, but now *each component is conserved separately*, a direct use of the vector methods of Chapter 3. We write conservation of momentum once for the x-direction and once for the y-direction and solve.

Example 9.5: A two-dimensional crash

A $1500\ \text{kg}$ car traveling east at $25.0\ \text{m/s}$ collides at an intersection with a $2500\ \text{kg}$ truck traveling north at $20.0\ \text{m/s}$. The two stick together. Find the magnitude and direction of the wreckage's velocity.



•Figure description
A vector diagram for a two-dimensional collision, with one momentum component pointing east and another north. They add head-to-tail to give the total momentum of the wreckage at an angle theta.

Solution

Step 1: Momentum before, one component at a time. Momentum is a vector, so it is conserved separately in each direction. Take east as $+x$ and north as $+y$. Before the collision the car carries all the eastward momentum (the truck has no east–west motion) and the truck carries all the northward momentum:

$$\begin{aligned} p_x &= (1500\,\text{kg})(25.0\,\text{m/s}) = 3.75 \times 10^4\,\text{kg}\cdot\text{m/s}, \\ p_y &= (2500\,\text{kg})(20.0\,\text{m/s}) = 5.00 \times 10^4\,\text{kg}\cdot\text{m/s}. \end{aligned} \tag{9.4.10}$$

Step 2: Momentum after. The wreckage has the combined mass $M = 1500 + 2500 = 4000\ \text{kg}$ and carries both components unchanged. Since $p_x = Mv_x$, divide by M to get each velocity component:

$$\begin{aligned} v_x &= \frac{p_x}{M} = \frac{3.75 \times 10^4}{4000}\,\text{m/s} = 9.375\,\text{m/s}, \\ v_y &= \frac{p_y}{M} = \frac{5.00 \times 10^4}{4000}\,\text{m/s} = 12.5\,\text{m/s}. \end{aligned} \tag{9.4.11}$$

Step 3: Speed and direction. The two components are the legs of a right triangle:

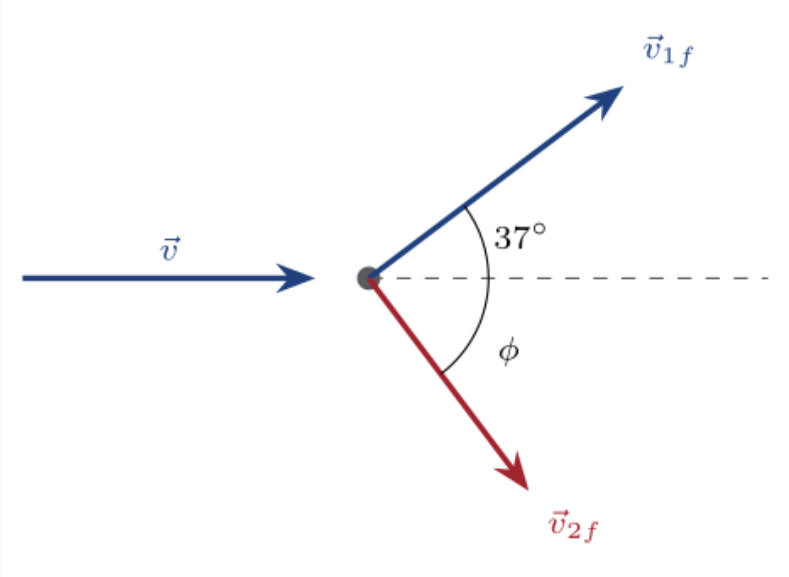
$$v = \sqrt{v_x^2 + v_y^2} = \sqrt{(9.375)^2 + (12.5)^2}\,\text{m/s} = \sqrt{244.1}\,\text{m/s} \approx \boxed{15.6\,\text{m/s}}, \tag{9.4.12}$$

$$\theta = \arctan\frac{v_y}{v_x} = \arctan\frac{12.5}{9.375} = \arctan(1.333) \approx \boxed{53.1^\circ\ \text{north of east}}. \tag{9.4.13}$$

Both components are positive, so the direction is in the first quadrant (north of east) and the calculator's angle needs no correction. Because M cancels in $v_y/v_x = p_y/p_x$, the direction depends only on the two momenta: the truck brought more momentum, so the wreckage heads more north than east.

Example 9.6: A glancing collision of equal masses

A proton moving at $3.50 \times 10^5\ \text{m/s}$ makes a glancing elastic collision with a second proton at rest. Afterward the first proton moves off at 37° from its original direction and the second at an angle ϕ on the other side. Find ϕ and the two final speeds.



•Figure description
A glancing collision seen from above. One ball moves off at 37 degrees to the original direction and the second at an angle phi on the opposite side, with final velocity vectors labeled.

Solution

Why the paths are perpendicular. Momentum conservation, with the target at rest and the equal masses canceled, reads $\mathbf{v} = \mathbf{v}_{1f} + \mathbf{v}_{2f}$. Take the dot product of each side with itself (recall $\mathbf{A} \cdot \mathbf{A} = A^2$ from Chapter 3):

$$v^2 = (\mathbf{v}_{1f} + \mathbf{v}_{2f}) \cdot (\mathbf{v}_{1f} + \mathbf{v}_{2f}) = v_{1f}^2 + v_{2f}^2 + 2\,\mathbf{v}_{1f} \cdot \mathbf{v}_{2f}. \tag{9.4.14}$$

Energy conservation, with $\frac{1}{2}m$ canceled, says $v^2 = v_{1f}^2 + v_{2f}^2$. Subtracting the two equations leaves $2\ \mathbf{v}_{1f} \cdot \mathbf{v}_{2f} = 0$. A zero dot product between two nonzero vectors means they are perpendicular. (The only other possibility is that one final velocity is zero: a head-on hit, in which the first proton stops dead.) Hence

$$\phi = 90^\circ - 37^\circ = \boxed{53^\circ}. \tag{9.4.15}$$

The final speeds. Because the two outgoing paths are perpendicular, they can serve as a pair of axes. Resolving the initial momentum $m\mathbf{v}$ along each of them, and remembering that the target proton brings no momentum to the collision, gives $mv_{1f} = mv\cos 37^\circ$ along the first path and $mv_{2f} = mv\cos 53^\circ$ along the second. (Each outgoing path makes the stated angle with the original direction, so the initial momentum's component along it is the adjacent side, the cosine.) The mass cancels from both:

$$v_{1f} = v\cos 37^\circ = (3.50 \times 10^5\ \mathrm{m/s})(0.799) \approx \boxed{2.80 \times 10^5\ \mathrm{m/s}}, \tag{9.4.16}$$

$$v_{2f} = v\cos 53^\circ = (3.50 \times 10^5\ \mathrm{m/s})(0.602) \approx \boxed{2.11 \times 10^5\ \mathrm{m/s}}. \tag{9.4.17}$$

Check. Energy requires $v_{1f}^2 + v_{2f}^2 = v^2$. Since $53^\circ = 90^\circ - 37^\circ$, $\cos 53^\circ = \sin 37^\circ$, so $v_{1f}^2 + v_{2f}^2 = v^2(\cos^2 37^\circ + \sin^2 37^\circ) = v^2$. ✓ Numerically, with the unrounded speeds, $(2.795)^2 + (2.106)^2 = 7.81 + 4.44 = 12.25$, and $(3.50)^2 = 12.25$ (both in units of $10^{10}\ \mathrm{m^2/s^2}$).

Checkpoint 9.3

A cart moving at speed v on a frictionless track strikes an identical cart at rest. What is each cart's velocity afterward if (a) they stick together, and (b) the collision is elastic? In which case is kinetic energy lost, and what fraction?

Answer

Answers to Checkpoints

Checkpoint 9.3. (a) Perfectly inelastic: momentum conservation gives $mv = (2m)v_f$, so both carts move off together at $v/2$. (b) Elastic: with equal masses, equations (9.11) give $v_{1f} = 0$ and $v_{2f} = v$; the moving cart stops dead and the struck cart carries on at the full speed, as in Newton's cradle. Kinetic energy is lost only in case (a): it drops from $\frac{1}{2}mv^2$ to $\frac{1}{2}(2m)(v/2)^2 = \frac{1}{4}mv^2$, so half of it is lost (the fraction $m_1/(m_1 + m_2) = 1/2$). In case (b) none is lost.

This page titled 9.4: Collisions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.5: Center of Mass

So far we have spoken of “an object” as though it were a single point. Real bodies are extended, and systems may contain many parts moving differently. The idea that tames this complexity is the **center of mass**: the mass-weighted average position of all the matter in a system. For a set of particles of masses m_i at positions $\mathbf{r}_i$, with total mass $M = \sum m_i$,

Key Equation: Center of mass

$$\mathbf{r}_{\mathrm{cm}} = \frac{1}{M}\sum_i m_i \mathbf{r}_i, \tag{9.12}$$

which in components reads $x_{\mathrm{cm}} = \frac{1}{M}\sum m_i x_i$, and similarly for y and z. Heavier particles pull the center of mass toward themselves; for a symmetric, uniform body it sits at the geometric center.

Example 9.7: Center of mass of three particles

Three particles lie in the xy-plane: $2.0\ \mathrm{kg}$ at the origin, $3.0\ \mathrm{kg}$ at $(4.0, 0)$ m, and $5.0\ \mathrm{kg}$ at $(0, 6.0)$ m. Where is the center of mass?

Solution

The total mass is $M = 2.0 + 3.0 + 5.0 = 10.0\ \mathrm{kg}$. Applying equation (9.12) component by component:

$$x_{\mathrm{cm}} = \frac{(2.0)(0) + (3.0)(4.0) + (5.0)(0)}{10.0} = \frac{12}{10} = 1.2\,\mathrm{m}, \tag{9.5.1}$$

$$y_{\mathrm{cm}} = \frac{(2.0)(0) + (3.0)(0) + (5.0)(6.0)}{10.0} = \frac{30}{10} = 3.0\,\mathrm{m}. \tag{9.5.2}$$

The center of mass is at $\boxed{(1.2,\ 3.0)\,\mathrm{m}}$: pulled toward the heavy $5.0\ \mathrm{kg}$ particle, and at no particle’s location.

Motion of the center of mass

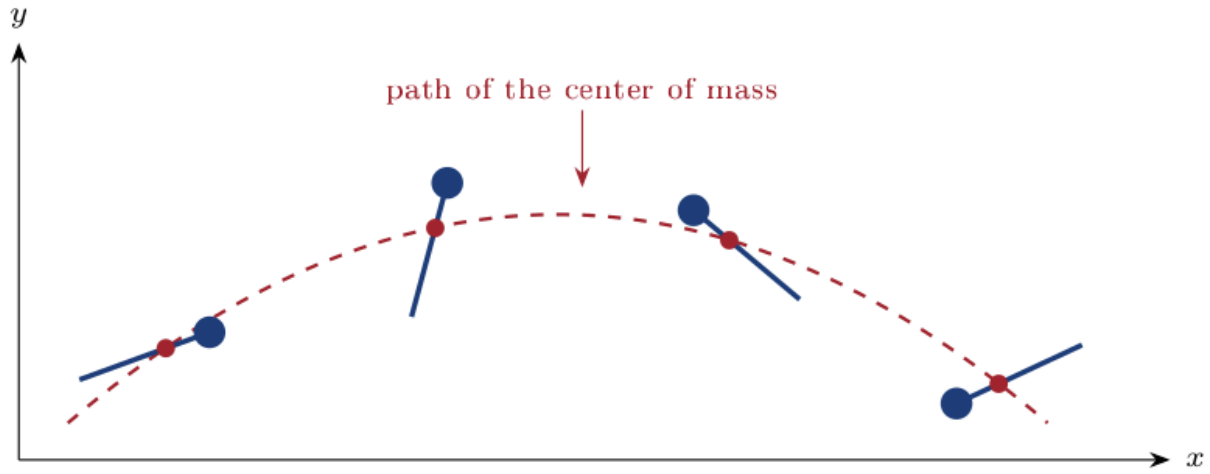


Figure 9.3. A wrench thrown across a room tumbles about a complicated path, but its center of mass (red) follows the same simple parabola as a point mass thrown the same way. The internal forces that hold the wrench together, and that make it spin, cannot shift the center of mass; only gravity, the external force, can.

•Figure description

A wrench thrown across a room, drawn at successive instants as it tumbles end over end. Its center of mass, marked in red and lying closer to the heavy end than to the middle, traces a smooth parabola even though the rest of the wrench follows a complicated path.

The center of mass is far more than a geometric curiosity. Differentiating equation (9.12) once shows that the total momentum of a system is $\mathbf{P} = M\mathbf{v}_{\text{cm}}$; differentiating a second time shows that the center of mass obeys Newton's second law in terms of the *external* forces alone:

Key Equation: Newton's second law for a system

$$\mathbf{F}_{\text{ext,net}} = M\mathbf{a}_{\text{cm}}. \tag{9.13}$$

Internal forces (the explosions, collisions, and pushes between the parts) never affect the motion of the center of mass. A wrench tumbling through the air spins wildly, but its center of mass traces the same simple parabola as a thrown ball (Figure 9.3). A shell bursting in flight scatters fragments everywhere, yet their center of mass continues along the original trajectory until a fragment hits the ground. And equation (9.13) contains the conservation law of §9.3 as a special case: when no external force acts, $\mathbf{a}_{\text{cm}} = 0$ and the center of mass glides at constant velocity, which is just another way of saying the total momentum is conserved.

Checkpoint 9.4

An astronaut floating at rest in deep space throws a wrench. Describe what happens to the astronaut, and explain why the center of mass of the astronaut-plus-wrench system stays exactly where it was.

Answer

Answers to Checkpoints

Checkpoint 9.4. The astronaut recoils in the direction opposite the throw. The total momentum starts at zero and no external force acts, so it stays zero: the wrench's momentum and the astronaut's are equal and opposite, and the astronaut moves more slowly in proportion to having more mass. Because the total momentum is zero, the velocity of the center of mass, $\mathbf{v}_{\text{cm}} = \mathbf{P}/M$, is zero before and after the throw. The center of mass was at rest and stays exactly where it was, even as the wrench and the astronaut move apart on either side of it.

This page titled 9.5: Center of Mass is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.S: Linear Momentum and Center of Mass (Summary)

Summary

- **Linear momentum** is $\mathbf{p} = m\mathbf{v}$, a vector. Newton's second law in its general form is $\mathbf{F}_{\text{net}} = d\mathbf{p}/dt$, reducing to $m\mathbf{a}$ when the mass is constant.
- The **impulse–momentum theorem**: $\mathbf{J} = \int \mathbf{F}\ dt = \mathbf{F}_{\text{avg}}\Delta t = \Delta\mathbf{p}$. For a fixed momentum change, a longer contact time means a smaller force: the principle behind airbags and cushioned catches.
- **Conservation of momentum**: the total momentum of a system is $\mathbf{P} = \sum_i \mathbf{p}_i$. When the net external force is zero, $\mathbf{P}$ is constant. Internal forces cancel in pairs and cannot change it.
- In all **collisions** momentum is conserved. *Elastic* collisions also conserve kinetic energy; *inelastic* ones do not; *perfectly inelastic* ones stick together. In two dimensions, conserve each momentum component separately.
- The **center of mass** is the mass-weighted average position, $\mathbf{r}_{\text{cm}} = \frac{1}{M}\sum m_i\mathbf{r}_i$. It carries the system's total momentum, $\mathbf{P} = M\mathbf{v}_{\text{cm}}$, and moves according to $\mathbf{F}_{\text{ext,net}} = M\mathbf{a}_{\text{cm}}$, unaffected by internal forces.

Key Equations

Linear momentum

$$\mathbf{p} = m\mathbf{v} \tag{9.S.1}$$

Newton's second law (momentum form)

$$\mathbf{F}_{\text{net}} = \frac{d\mathbf{p}}{dt} \tag{9.S.2}$$

Impulse–momentum theorem

$$\mathbf{J} = \int \mathbf{F}\, dt = \mathbf{F}_{\text{avg}}\, \Delta t = \Delta\mathbf{p} \tag{9.S.3}$$

Conservation of linear momentum

$$\text{if } \mathbf{F}_{\text{ext,net}} = 0, \qquad \mathbf{P}_i = \mathbf{P}_f \quad (\text{total momentum constant}) \tag{9.S.4}$$

Elastic collisions in one dimension

$$\begin{aligned} v_{1f} &= \frac{m_1 - m_2}{m_1 + m_2}\, v_{1i} + \frac{2m_2}{m_1 + m_2}\, v_{2i}, \\ v_{2f} &= \frac{2m_1}{m_1 + m_2}\, v_{1i} + \frac{m_2 - m_1}{m_1 + m_2}\, v_{2i} \end{aligned} \tag{9.S.5}$$

Center of mass

$$\mathbf{r}_{\text{cm}} = \frac{1}{M}\sum_i m_i\mathbf{r}_i \tag{9.S.6}$$

Newton's second law for a system

$$\mathbf{F}_{\text{ext,net}} = M\mathbf{a}_{\text{cm}} \tag{9.S.7}$$

This page titled 9.S: Linear Momentum and Center of Mass (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

9.E: Linear Momentum and Center of Mass (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 9.1 — Linear Momentum

9.1 ★ Find the magnitude of the momentum of (a) a $1400\ \text{kg}$ car at $30\ \text{m/s}$, and (b) a $0.045\ \text{kg}$ golf ball at $70\ \text{m/s}$.

9.2 ★ *[Conceptual.]* A bowling ball and a ping-pong ball have the same momentum.

a. Which is moving faster, and why?
b. Which has more kinetic energy? *(Hint: express K in terms of p and m.)*

9.3 ★★ A $0.16\ \text{kg}$ hockey puck moving east at $18\ \text{m/s}$ is struck and afterward moves north at $22\ \text{m/s}$.

a. Find the change in the puck's momentum (magnitude and direction).
b. What impulse did the stick deliver to the puck, and in what direction did it point?

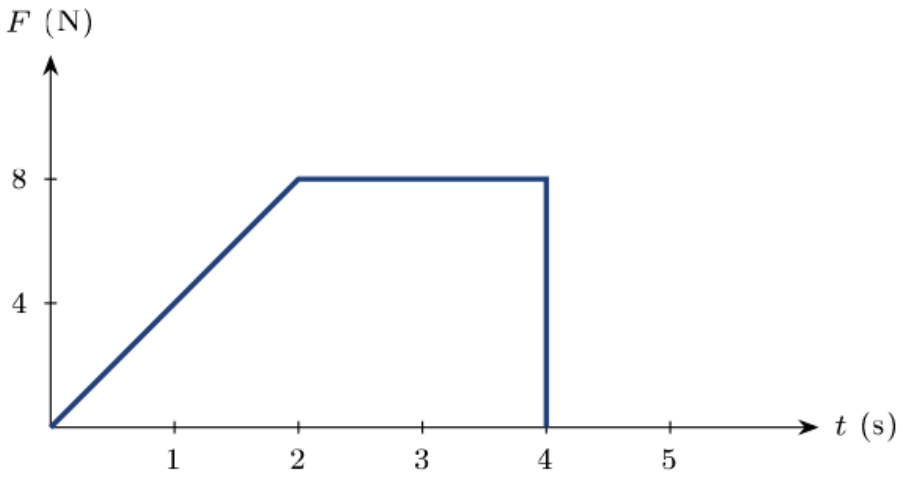
Section 9.2 — Impulse

9.4 ★ A $0.145\ \text{kg}$ baseball arrives at $40\ \text{m/s}$ and leaves the bat at $50\ \text{m/s}$ in the opposite direction. The bat is in contact with the ball for $1.2\ \text{ms}$.

a. Find the impulse delivered to the ball.
b. Find the average force exerted by the bat.

9.5 ★ *[Conceptual.]* Explain, in terms of the impulse–momentum theorem, why each of the following reduces the force of an impact: (a) an airbag, (b) bending your knees when you land from a jump, (c) a boxer "rolling" with a punch.

9.6 ★★ *[Graph-reading.]* The graph shows the net force on a $2.0\ \text{kg}$ cart (initially at rest) as a function of time.



•Figure description
A graph of net force in newtons versus time in seconds acting on a two kilogram cart that starts at rest; the area under the graph gives the impulse.

a. Find the total impulse delivered to the cart (the area under the graph).
b. Find the cart's speed at $t = 4\ \text{s}$.

9.7 ★★ *[Real-world: safety.]* A $75\ \text{kg}$ driver in a $14\ \text{m/s}$ crash is brought to rest. Compare the average force on the driver if the stop takes $0.20\ \text{s}$ (with an airbag) versus $0.020\ \text{s}$ (hitting a rigid dashboard).

9.8 ★★★ A $0.50\ \text{kg}$ ball strikes a wall moving horizontally at $12\ \text{m/s}$ and rebounds at $9.0\ \text{m/s}$. The contact lasts $8.0\ \text{ms}$.

a. Find the impulse the wall delivers to the ball.
b. Find the average force on the ball.
c. By Newton's third law, what average force does the ball exert on the wall?

Section 9.3 — Conservation of Linear Momentum

9.9 ★ A $4.0\ \text{kg}$ rifle fires a $0.020\ \text{kg}$ bullet at $350\ \text{m/s}$. Find the recoil speed of the rifle.

9.10 ★★ *[Real-world: spaceflight.]* A spacecraft of total mass $1200\ \text{kg}$ (propellant included) drifts at rest in deep space. It ejects $15\ \text{kg}$ of exhaust gas at $2500\ \text{m/s}$, measured in the frame in which the spacecraft was originally at rest. Find the speed the spacecraft acquires, and explain why this works in a vacuum, with nothing to push against.

9.11 ★★ A $60\ \text{kg}$ skater and a $40\ \text{kg}$ skater stand at rest on frictionless ice and push off each other. If the $60\ \text{kg}$ skater moves away at $2.0\ \text{m/s}$, how fast does the other move, and in which direction?

Section 9.4 — Collisions

9.12 ★ *[Conceptual.]* For each collision, state whether kinetic energy is conserved and whether momentum is conserved: (a) two balls of clay that stick together, (b) two steel balls that bounce apart with no loss of speed, (c) a car crash in which the cars crumple but bounce apart.

9.13 ★★ A $0.50\ \text{kg}$ cart moving at $3.0\ \text{m/s}$ collides with a stationary $1.5\ \text{kg}$ cart and they stick together. Find (a) their common speed afterward, and (b) the kinetic energy lost in the collision.

9.14 ★★ A $2.0\ \text{kg}$ ball moving right at $5.0\ \text{m/s}$ collides elastically with a stationary $3.0\ \text{kg}$ ball. Use the elastic-collision relations to find the velocity of each ball afterward.

9.15 ★★★ *[Real-world: forensics.]* A $1200\ \text{kg}$ car heading east at an unknown speed runs a light and strikes a $1600\ \text{kg}$ car heading north at $16\ \text{m/s}$. The wreckage slides off together at 35° north of east. Reconstruct the eastbound car's speed before impact.

9.16 ★★★ A ballistic pendulum: a $12\ \text{g}$ bullet embeds in a $2.0\ \text{kg}$ block hanging from a string. The block (with bullet) then swings up to a height of $0.18\ \text{m}$.

a. Find the speed of the block-plus-bullet just after impact (use energy conservation for the swing).
b. Find the bullet's speed just before impact (use momentum conservation for the collision).
c. Why must you use momentum for one stage and energy for the other?

Section 9.5 — Center of Mass

9.17 ★ Two particles lie on the x-axis: $3.0\ \text{kg}$ at $x = 0\ \text{m}$ and $1.0\ \text{kg}$ at $x = 2.0\ \text{m}$. Find the center of mass.

9.18 ★★ *[Real-world: astronomy.]* The Earth ($5.97 \times 10^{24}\ \text{kg}$) and the Moon ($7.35 \times 10^{22}\ \text{kg}$) are separated by $3.84 \times 10^{8}\ \text{m}$. How far from the Earth's center is the center of mass of the Earth–Moon system? Compare this with the Earth's radius ($6.37 \times 10^{6}\ \text{m}$).

9.19 ★★ A $72\ \text{kg}$ person stands at one end of a stationary $38\ \text{kg}$ canoe, $4.5\ \text{m}$ long, floating on still water. The person walks to the other end. Treat the water as frictionless.

a. Where is the center of mass of the person-plus-canoe system while the person walks? Explain using conservation of momentum.
b. Which way, and how far, does the canoe move?
c. How far does the person move relative to the water?

9.20 ★★★ *[Conceptual.]* A fireworks shell is launched and follows a parabolic path; at the top of its arc it explodes into many fragments.

a. Describe the path of the center of mass of the fragments immediately after the explosion (before any fragment lands). Justify your answer.
b. Does the explosion (an internal force) change the motion of the center of mass? Explain.

Challenge Problems

9.21 ★★★ A $0.20\ \text{kg}$ puck moving at $4.0\ \text{m/s}$ strikes an identical stationary puck on frictionless ice in a glancing *elastic* collision. Afterward the incoming puck is deflected 30° from its original direction.

a. Show that, for equal masses with one initially at rest, an elastic collision sends the two pucks off at 90° to each other.
b. Find the speed of each puck after the collision.
c. At what angle does the struck puck move?
d. Verify that kinetic energy is conserved.

9.22 ★★★ A $6.0\ \text{kg}$ object initially at rest explodes into three $2.0\ \text{kg}$ fragments. Fragment A flies east at $30\ \text{m/s}$ and fragment B flies north at $30\ \text{m/s}$.

a. Find the momentum vectors of fragments A and B.
b. Use momentum conservation to find the momentum of fragment C.
c. Find the speed and direction of fragment C.
d. How much kinetic energy was released in the explosion?

9.23 ★★★ *[Real-world: spaceflight.]* A rocket in deep space (no gravity) ejects exhaust at a constant speed v_{ex} relative to itself.

a. By applying momentum conservation to the rocket and a small ejected mass, show that $m\ dv = -v_{\text{ex}}\ dm$, and integrate to obtain $\Delta v = v_{\text{ex}} \ln(m_i/m_f)$.
b. A rocket with $v_{\text{ex}} = 2500\ \text{m/s}$ and an initial mass of $5000\ \text{kg}$ must gain $\Delta v = 4500\ \text{m/s}$. Find the final mass and the mass of propellant burned.
c. What fraction of the initial mass is propellant?
d. Explain why a multistage rocket can reach a larger Δv than a single stage carrying the same total propellant.

This page titled 9.E: Linear Momentum and Center of Mass (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10: ROTATION I

Every motion we have studied so far has been *translation*: objects sliding, falling, or flying from one place to another, treated as single points. But the world is full of *rotation*: wheels and gears, spinning drills, opening doors, the Earth turning on its axis. A rigid body that rotates has every particle moving at once, each on its own circle, yet the whole turns together as a unit. Describing that requires a new vocabulary.

The good news is that rotation mirrors translation almost exactly. For every linear quantity there is an angular partner (position becomes angle, velocity becomes angular velocity, mass becomes *rotational inertia*, force becomes *torque*), and every law we have learned reappears in rotational form. This chapter builds the first half of that dictionary. It describes rotation, ties it to the forward motion of a rolling wheel, measures how strongly a body resists being turned, and arrives at Newton's second law for rotation, with torque given its full vector definition through the right-hand rule. Chapter 11 completes the dictionary with rotational kinetic energy, work, and angular momentum.

Learning Objectives

- Describe rotation using angular displacement, velocity, and acceleration, and relate them to linear quantities.
- Apply the rotational kinematic equations for constant angular acceleration.
- Relate the forward motion of a body rolling without slipping to its rotation, and find the velocity of any point on it.
- Compute the rotational inertia (moment of inertia) of a system of particles or a continuous body.
- Find torque as a magnitude and as a vector, using the cross product and the right-hand rule.
- Apply Newton's second law for rotation, $\sum \tau = I\alpha$, including to a body rolling down a ramp.

10.1: Rotational Kinematics

10.2: Rolling Motion

10.3: Rotational Inertia

10.4: Torque and the Right-Hand Rule

10.S: Rotation I (Summary)

10.E: Rotation I (Problems)

This page titled 10: Rotation I is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.1: Rotational Kinematics

To measure rotation we measure *angle*. The natural unit is the **radian**: the angle subtended at the center of a circle by an arc equal in length to the radius. For an arc of length s on a circle of radius r,

$$\theta = \frac{s}{r}, \tag{10.1}$$

so the radian is a ratio of two lengths: dimensionless. One full circle is

$$1 \text{ rev} = 2\pi \text{ rad} = 360^\circ, \tag{10.1.1}$$

and that single statement is the conversion factor among all three units.

Note: Radians are not optional

Because the radian is dimensionless it quietly drops out of unit bookkeeping, which makes it easy to forget, and yet every relation linking a linear quantity to an angular one ($s = r\theta$, $v = r\omega$, $a_t = r\alpha$, $a_c = r\omega^2$, $W = \tau\ \Delta\theta$) holds *only* when the angle is measured in radians. A wheel of radius $0.20\ \text{m}$ turned through 30° does not sweep an arc of $(0.20)(30) = 6.0\ \text{m}$. Convert first: $30^\circ = 0.52\ \text{rad}$, and the arc is $0.10\ \text{m}$.

Degrees and revolutions remain perfectly good units for *describing* a rotation, and rpm is the everyday unit for machinery. Use whichever unit the problem hands you while you are working with angular quantities alone, but the moment a radius enters the calculation, convert to radians.

The **angular displacement** is the change in angle, $\Delta\theta = \theta - \theta_0$; the **angular velocity** $\omega = d\theta/dt$ (rad/s) is its rate of change; and the **angular acceleration** $\alpha = d\omega/dt$ (rad/s^2) is the rate of change of that. These are exactly parallel to Δx, v, and a in straight-line motion.

Rotation rates are often quoted not as ω but as a **frequency**: a count of revolutions per unit time. If the body completes one revolution in a **period** T, then its frequency is $f = 1/T$, and since one revolution is 2π radians,

$$\omega = 2\pi f = \frac{2\pi}{T}. \tag{10.2}$$

Machinery is nearly always rated in revolutions per minute. To convert, divide by 60 to get rev/s, then multiply by 2π: a motor at $1800\ \text{rpm}$ turns at $(1800/60)(2\pi) = 188\ \text{rad/s}$.

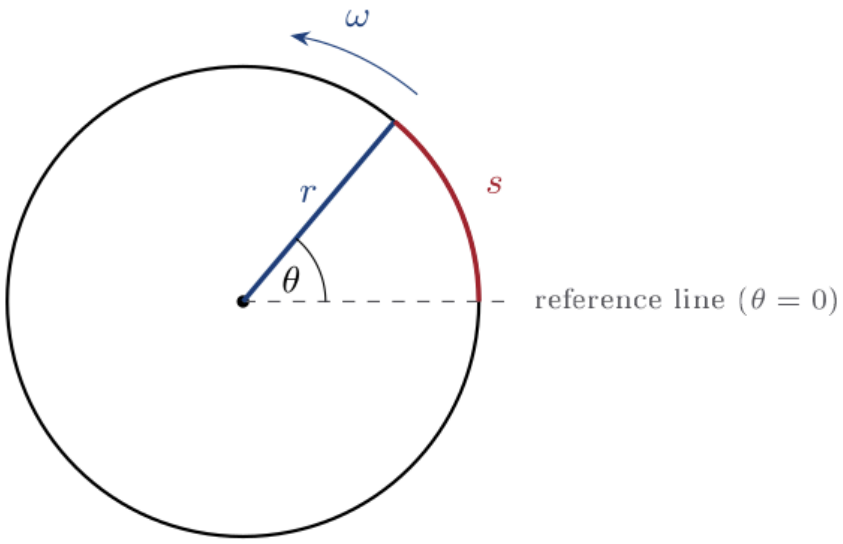


Figure 10.1. Angular position is measured from a reference line. A point at radius r moves along an arc $s = r\theta$ as the body turns through angle θ ; the angular velocity ω gives the rate of turning.

•Figure description

A rotating disc with a reference line marking angular position. A point at radius r sweeps along an arc of length s equal to r theta as the body turns through angle theta, with omega marking the direction of rotation.

Each particle of a rotating body also has an ordinary linear speed along its circular path, tied to the angular quantities by the radius:

Key Equation: Relating linear and angular motion

$$s = r\theta, \qquad v = r\omega, \qquad a_t = r\alpha, \qquad a_c = r\omega^2 = \frac{v^2}{r}. \tag{10.3}$$

Here v is the tangential (linear) speed of a point at radius r, and a_t is its *tangential* acceleration: the part that changes its speed. Points farther from the axis move faster because every point shares the same ω and α, so a larger r means a larger v; this is why the rim of a wheel travels faster than a point near its hub (Figure 10.1).

A point on a rotating body is also moving in a circle, so it has the **centripetal acceleration** $a_c = v^2/r$ of Chapter 4, directed inward along the radius. Substituting $v = r\omega$ puts it in convenient angular form, $a_c = r\omega^2$. The two components are perpendicular to each other, so the magnitude of the total linear acceleration of the point is

$$a = \sqrt{a_t^2 + a_c^2}. \tag{10.4}$$

The two do very different jobs: a_t changes how fast the point is going, while a_c changes only the direction it is going. A body spinning at a *constant* rate has $\alpha = 0$ and therefore $a_t = 0$, yet every point on it still has $a_c \neq 0$. Only at an instant when $\omega = 0$ (a body starting from rest, say) does a_c vanish and the total acceleration reduce to $r\alpha$ alone.

Checkpoint 10.1

A hard-drive platter starts from rest and spins up at a constant angular acceleration.

a. At the instant it begins to turn, what are a_t and a_c for a point on its rim?
b. A few seconds later, which of the two is still growing, and which is not? Explain using $a_t = r\alpha$ and $a_c = r\omega^2$.
c. Once the platter reaches its operating speed and holds it steady, what are a_t and a_c then?

Answer

Because the definitions of ω and α match those of v and a, the kinematic equations of Chapter 2 carry over unchanged, with angular symbols in place of linear ones. As before, they hold only for *constant* angular acceleration:

Key Equation: Rotational kinematic equations (constant α)

$$\omega = \omega_0 + \alpha\, t, \tag{10.5}$$
$$\Delta\theta = \omega_0\, t + \tfrac{1}{2}\alpha\, t^2, \tag{10.6}$$
$$\omega^2 = \omega_0^2 + 2\alpha\, \Delta\theta, \tag{10.7}$$
$$\Delta\theta = \tfrac{1}{2}(\omega + \omega_0)\, t. \tag{10.8}$$

Set these beside equations (2.8)–(2.11) and the correspondence is exact: the same four equations, with θ, ω, and α written in place of x, v, and a. Everything you already know about constant-acceleration problems (which equation to reach for, how to keep the signs straight) carries over without modification.

Example 10.1: Revving a crankshaft

An engine crankshaft idling at $1000\ \text{rpm}$ revs up to $5200\ \text{rpm}$ in $1.2\ \text{s}$ under constant angular acceleration. Find that acceleration in (a) rev/min^2 and (b) deg/s^2.

Solution

The angular acceleration is $\alpha = \Delta\omega/\Delta t$. The change in angular velocity is final minus initial, $\Delta\omega = 5200 - 1000 = 4200\ \text{rpm}$. The whole problem is then a unit conversion; the only rule is that the time unit in $\Delta\omega$ and the time unit in Δt must be handled deliberately.

(a) Keep rev/min, so the time must also be in minutes. Since $1\ \text{min} = 60\ \text{s}$,

$$\Delta t = 1.2\ \cancel{\text{s}} \times \frac{1\,\text{min}}{60\ \cancel{\text{s}}} = 0.020\,\text{min}. \tag{10.1.2}$$

Then

$$\alpha = \frac{4200\,\text{rev/min}}{0.020\,\text{min}} = \boxed{2.1 \times 10^5\,\text{rev/min}^2}. \tag{10.1.3}$$

Dividing (rev/min) by min gives rev/min^2: one “per minute” from the angular velocity and one from the time.

(b) Now convert $\Delta\omega$ to deg/s with a chain of conversion factors, one revolution being 360°:

$$4200\ \frac{\cancel{\text{rev}}}{\cancel{\text{min}}} \times \frac{360^\circ}{1\ \cancel{\text{rev}}} \times \frac{1\ \cancel{\text{min}}}{60\,\text{s}} = \frac{(4200)(360)}{60}\ \frac{\text{deg}}{\text{s}} = 25200\,\text{deg/s}. \tag{10.1.4}$$

Then

$$\alpha = \frac{25200\,\text{deg/s}}{1.2\,\text{s}} = \boxed{2.1 \times 10^4\,\text{deg/s}^2}. \tag{10.1.5}$$

Check. Convert the answer to (a) directly: $2.1 \times 10^5\ \frac{\text{rev}}{\text{min}^2} \times \frac{360^\circ}{\text{rev}} \times \left(\frac{1\,\text{min}}{60\,\text{s}}\right)^2 = \frac{(2.1\times10^5)(360)}{3600}\ \frac{\text{deg}}{\text{s}^2} = 2.1 \times 10^4\ \text{deg/s}^2$. The minutes are s*quared*, so the time factor must be squared too, the same rule as for areas and volumes in Chapter 1.

Answers to Checkpoints

Checkpoint 10.1. (a) At the first instant $\omega = 0$, so $a_c = r\omega^2 = 0$, while $a_t = r\alpha$ already has its full, constant value. (b) The tangential acceleration $a_t = r\alpha$ is not growing: α is constant. The centripetal acceleration $a_c = r\omega^2$ is growing, because ω increases steadily, and since ω is squared, a_c grows ever faster. (c) At steady operating speed $\alpha = 0$, so $a_t = 0$, while $a_c = r\omega^2$ is constant and large. The rim still accelerates, but only toward the center.

This page titled 10.1: Rotational Kinematics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.2: Rolling Motion

Section 10.1 treated a body turning about a fixed axis. A wheel rolling along a road is different, because its axis moves too. A rolling wheel is doing two things at once: its center moves forward (translation) while the wheel turns about that center (rotation). The two are locked together by the condition of **rolling without slipping**. In one full turn the wheel advances exactly one circumference, $2\pi R$; for any angle θ, measured in radians, it advances an arc length $s = R\theta$. Differentiating ties the center-of-mass speed to the angular speed, and differentiating again ties the accelerations:

Key Equation: Rolling without slipping

$$x_{\mathrm{cm}} = R\theta, \qquad v_{\mathrm{cm}} = R\omega, \qquad a_{\mathrm{cm}} = R\alpha. \tag{10.9}$$

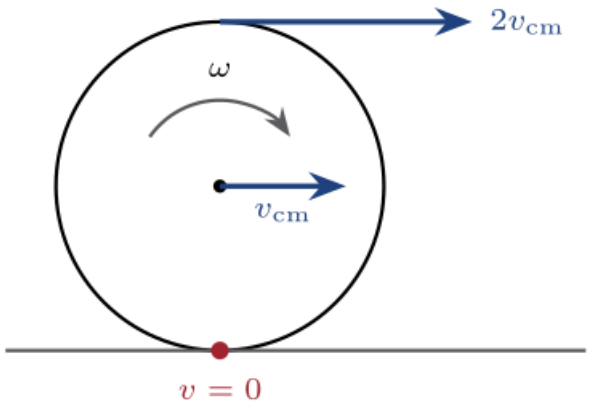


Figure 10.2. Rolling without slipping as translation plus rotation. The center moves at v_{cm}, the top at $2v_{\mathrm{cm}}$, and the contact point at the bottom is instantaneously at rest.

•Figure description

Rolling without slipping shown as translation combined with rotation. The center of the wheel moves at v sub cm, the top point at twice v sub cm, and the contact point at the ground is instantaneously at rest.

A useful way to picture rolling is to add the two motions point by point (Figure 10.2). The pure translation gives every point the velocity v_{cm}; the pure rotation gives the top of the wheel a forward velocity $R\omega = v_{\mathrm{cm}}$ and the bottom a backward velocity of the same size. Adding them, the top of the wheel moves at $2v_{\mathrm{cm}}$ while the contact point at the bottom is, for that instant, *at rest*, which is exactly why a rolling wheel does not skid.

Rolling without slipping is a purely kinematic condition: it says how a rolling body's motion is shared between translation and rotation, but not what makes the body roll or how much energy it carries. Both questions need ideas we have not met yet. Section 10.4 uses the moment of inertia and torque to find how fast a body rolls down a ramp, and Section 11.1 in Chapter 11 finds the kinetic energy of a rolling body.

Checkpoint 10.2

A car drives past you at $25\ \mathrm{m/s}$ without its tires slipping.

a. How fast is the top of a tire moving relative to the road?
b. How fast is the patch of tread touching the road moving relative to the road?
c. A point on the front edge of the tire, level with the axle, is moving at what speed, and in what direction?

Answer

Answers to Checkpoints

Checkpoint 10.2. Every point on the tire moves with the axle's forward velocity $v_{\text{cm}} = 25\ \text{m/s}$ plus a rotational velocity of size $\omega R = v_{\text{cm}}$ tangent to the tire. (a) At the top the two point the same way and add: $2v_{\text{cm}} = 50\ \text{m/s}$. (b) At the bottom they cancel exactly: the contact patch is momentarily at rest relative to the road; that is what "not slipping" means. (c) At the front, level with the axle, the rotational velocity points straight down (for a wheel rolling forward), perpendicular to the forward velocity. The two combine to $\sqrt{25^2 + 25^2} = 25\sqrt{2} \approx 35\ \text{m/s}$, directed 45° below the forward direction.

This page titled 10.2: Rolling Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.3: Rotational Inertia

In translation, mass measures an object's resistance to a change in motion. In rotation, that role is played by **rotational inertia**, or **moment of inertia** I, and it depends not just on how much mass there is but on *how that mass is distributed relative to the axis*. Mass far from the axis is much harder to spin up than mass close in. A single particle of mass m at distance r from the axis has $I = mr^2$; for an extended body we add up the contributions:

Key Equation: Moment of inertia

$$I = \sum_i m_i r_i^2 \quad \text{(particles)}, \qquad I = \int r^2\, dm \quad \text{(continuous body)}. \tag{10.10}$$

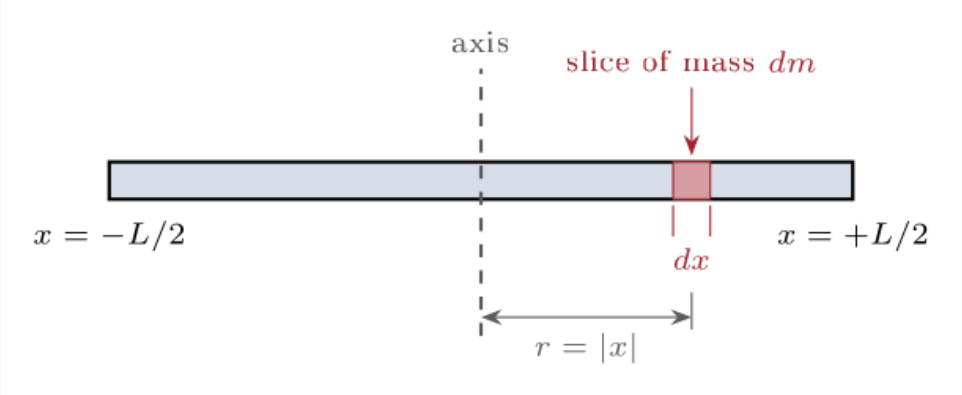
The r^2 weighting is why distribution matters so much: doubling a particle's distance from the axis quadruples its contribution. The SI unit of I is $\mathrm{kg \cdot m^2}$.

Example 10.2: Moment of inertia of a thin rod

Find the moment of inertia of a uniform thin rod of length L and mass M about an axis perpendicular to the rod through its center.

Solution

Step 1: Picture what dm means. The integral $I = \int r^2\ dm$ is an instruction: chop the body into tiny pieces, multiply each piece's mass dm by the square of its distance r from the axis, and add up all the pieces. For the rod, cut it into thin slices across its length, as in the figure. Measure position x along the rod from its center, so the rod runs from $x = -L/2$ to $x = +L/2$ and the axis sits at $x = 0$.



•Figure description
A thin rod lying along the x axis from minus L over 2 to plus L over 2, with a perpendicular axis of rotation through its center at x equals 0. A thin slice of the rod at position x has width d x and mass d m. Its distance from the axis is the absolute value of x.

Step 2: Describe each slice by its position. We cannot integrate "over mass" directly: there is no mass variable to put limits on. What we *can* describe is where each slice sits. So we rewrite everything inside the integral in terms of the single variable x, one piece at a time.

- *The distance.* A slice at position x is a distance $r = |x|$ from the axis, so $r^2 = x^2$ (squaring removes the absolute value).
- *The mass.* The slice has width dx. The rod is uniform, so every meter of it has the same mass, $\lambda = M/L$ (its linear mass density). A slice's mass is then (mass per length) $\times$ (length):

$$dm = \lambda\, dx = \frac{M}{L}\, dx. \tag{10.3.1}$$

- *The limits.* "Add up every piece of mass" now means "add up every slice from one end of the rod to the other": x runs from $-L/2$ to $+L/2$. The limits are positions, because the variable is now a position.

Two quick checks that the swap is right. Units: $(M/L)\ dx$ is $(\ \mathrm{kg/m})(\ \mathrm{m}) =\ \mathrm{kg}$, a mass, as dm must be. And adding up all the slices with no r^2 weighting must give back the whole rod:

$$\int dm = \int_{-L/2}^{L/2} \frac{M}{L}\,dx = \frac{M}{L}\Big[x\Big]_{-L/2}^{L/2} = \frac{M}{L}\left(\frac{L}{2}+\frac{L}{2}\right) = M. \checkmark \tag{10.3.2}$$

Now make all three replacements at once in the moment-of-inertia integral:

$$I = \int r^2\,dm \quad\longrightarrow\quad I = \int_{-L/2}^{L/2} x^2\,\frac{M}{L}\,dx = \frac{M}{L}\int_{-L/2}^{L/2} x^2\,dx. \tag{10.3.3}$$

The constant M/L comes outside the integral, leaving an ordinary power-rule integral in x.

Step 3: Integrate and evaluate. By the power rule, $\int x^2\ dx = x^3/3$. Evaluate at the upper limit and subtract the value at the lower limit. The cube of a negative number is negative, $(-L/2)^3 = -L^3/8$, so subtracting it *adds*:

$$\left[\frac{x^3}{3}\right]_{-L/2}^{L/2} = \frac{(L/2)^3}{3} - \frac{(-L/2)^3}{3} = \frac{L^3}{24} - \left(-\frac{L^3}{24}\right) = \frac{L^3}{24}+\frac{L^3}{24} = \frac{L^3}{12}. \tag{10.3.4}$$

Therefore

$$I = \frac{M}{L}\cdot\frac{L^3}{12} = \boxed{\tfrac{1}{12}ML^2}, \tag{10.3.5}$$

where one power of L canceled. The two halves of the rod contribute equally, which is why the answer is twice the $L^3/24$ from either half.

The end axis. If instead the axis passes through one *end*, x runs from 0 to L, and the same integral gives

$$I = \frac{M}{L}\int_0^L x^2\,dx = \frac{M}{L}\cdot\frac{L^3}{3} = \tfrac{1}{3}ML^2, \tag{10.3.6}$$

four times larger, because more of the mass now lies far from the axis. (The parallel-axis theorem checks this: $\frac{1}{12}ML^2 + M(L/2)^2 = \frac{1}{12}ML^2 + \frac{3}{12}ML^2 = \frac{1}{3}ML^2$.)

Integrals like this have been worked out for all the common shapes; Table 10.1 collects the most useful results. Each is some pure number times MR^2 or ML^2, with the number reflecting how the mass is spread.

Table 10.1. Moments of inertia for some uniform bodies of mass M. Axes pass through the center of symmetry unless noted otherwise; R is the radius and L the length.

Body (axis)	Moment of inertia
Thin hoop / ring (central axis)	MR^2
Solid cylinder or disk (central axis)	$\frac{1}{2}MR^2$
Thin rod (through center, $\perp$)	$\frac{1}{12}ML^2$
Thin rod (through end, $\perp$)	$\frac{1}{3}ML^2$
Solid sphere (through center)	$\frac{2}{5}MR^2$
Thin spherical shell (through center)	$\frac{2}{3}MR^2$

The parallel-axis theorem

Table 10.1 gives I about an axis through the center of mass, but real objects often turn about some other axis: a rod swinging

from one end, a door on its hinges, a pedal about a crank. Redoing the integral for every new axis would be tedious, and happily it is unnecessary. If you know I_{cm} about an axis through the center of mass, then about any *parallel* axis a distance d away,

Key Equation: Parallel-axis theorem

$$I = I_{\text{cm}} + Md^2. \tag{10.11}$$

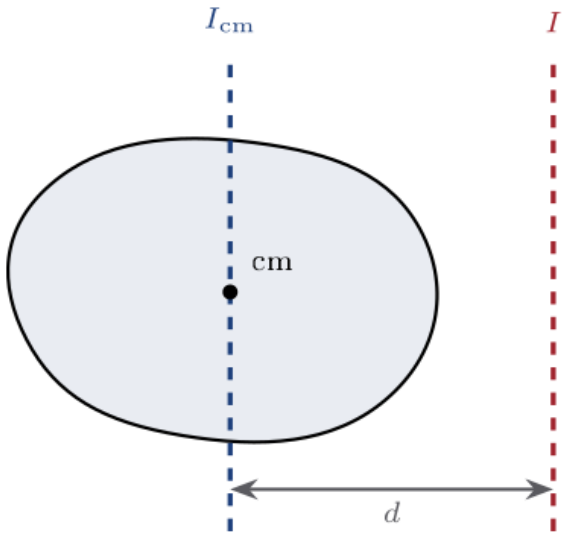


Figure 10.3. The parallel-axis theorem. If the moment of inertia about the blue axis through the center of mass is I_{cm}, then about any parallel axis a distance d away it is $I = I_{\text{cm}} + Md^2$.

•Figure description

The parallel-axis theorem. A body is shown with a blue axis through its center of mass and a second parallel axis a distance d away, about which the moment of inertia is larger by m d squared.

Here M is the total mass of the body and d is the perpendicular distance between the two axes (Figure 10.3). The two conditions are easy to state and easy to forget: the axes must be *parallel* to each other, and one of them must pass through the *center of mass*. The theorem says nothing about two arbitrary axes.

Two features of equation (10.11) are worth pausing on. First, Md^2 can never be negative, so *the moment of inertia is smallest about an axis through the center of mass*; every parallel axis is harder to spin the body about. Second, the added term is exactly what you would get by treating the entire body as a single point mass M sitting at the center of mass, a distance d from the axis.

The rod worked out above makes a ready check. Its center-of-mass value is $I_{\text{cm}} = \frac{1}{12}ML^2$, and the axis through one end is parallel to that one and a distance $d = L/2$ from it, so

$$I_{\text{end}} = \tfrac{1}{12}ML^2 + M\left(\frac{L}{2}\right)^2 = \tfrac{1}{12}ML^2 + \tfrac{1}{4}ML^2 = \tfrac{1}{3}ML^2, \tag{10.3.7}$$

which matches the integral exactly, and took one line instead of one integration.

Checkpoint 10.3

A solid disk of mass M and radius R has $I_{\text{cm}} = \frac{1}{2}MR^2$ about its central axis. Find its moment of inertia about a parallel axis through a point on its rim. Is the result larger or smaller than I_{cm}, and did it have to come out that way?

Answer

Answers to Checkpoints

Checkpoint 10.3. By the parallel-axis theorem with the axis a distance R from the center, $I = I_{\mathrm{cm}} + MR^2 = \frac{1}{2}MR^2 + MR^2 = \frac{3}{2}MR^2$. That is three times I_{cm}, and it had to be larger: the added term Md^2 is never negative, so the moment of inertia about the center of mass is the smallest of any parallel axis.

This page titled 10.3: Rotational Inertia is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.4: Torque and the Right-Hand Rule

What makes something rotate? Not just a force, but a force applied in the right place. Pushing a door near its hinges barely turns it; the same push at the handle swings it easily. The rotational effectiveness of a force is the **torque**. For a force F applied at a distance r from the axis, making an angle θ with the line from the axis to the point of application,

Key Equation: Torque

$$\tau = rF\sin\theta = r_\perp F, \tag{10.12}$$

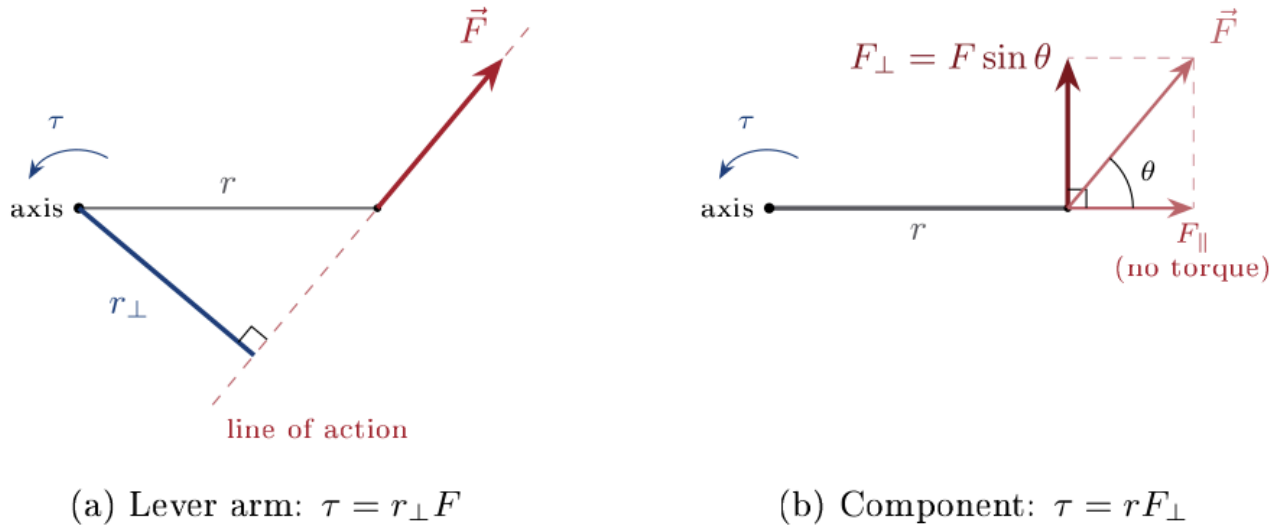


Figure 10.4. Two equivalent ways to find the torque $\tau = rF\sin\theta$ of a force $\mathbf{F}$ applied a distance r from the axis. (a) *Lever-arm view.* Keep the whole force and use the lever arm $r_\perp = r\sin\theta$, the perpendicular distance from the axis to the force's line of action: $\tau = r_\perp F$. (b) *Component view.* Keep the whole distance r and split the force. Only the component $F_\perp = F\sin\theta$, perpendicular to r, turns the body; $F_\parallel$ lies along the line through the axis and produces no torque: $\tau = rF_\perp$. Both give $rF\sin\theta$. Here the torque turns the body counterclockwise, which we take as positive.

•Figure description

Two panels showing the same force F applied at the same point, a distance r from a rotation axis, at angle theta to r. Panel a, the lever-arm view: the force's line of action is extended as a dashed line, and the lever arm r perpendicular is drawn from the axis to that line, meeting it at a right angle. The torque is r perpendicular times F. Panel b, the component view: the force is split into F perpendicular, at right angles to r, and F parallel, along r. Only F perpendicular turns the body; F parallel lies along the line through the axis. The torque is r times F perpendicular. In both panels the torque is counterclockwise.

where $r_\perp = r\sin\theta$ is the **lever arm**, the perpendicular distance from the axis to the line of the force (Figure 10.4a). Grouping the same two factors the other way gives an equally useful third form, $\tau = rF_\perp$, in which $F_\perp = F\sin\theta$ is the component of the force perpendicular to $\mathbf{r}$ (Figure 10.4b). Read either way, the message is identical: only the perpendicular part of the force does any turning. A force aimed straight away from the axis ($\theta = 0$) or straight at it ($\theta = 180^\circ$) has no perpendicular component and produces no torque at all. The SI unit is the newton-meter ($\mathrm{N}\cdot\mathrm{m}$). Torque has a sense (we will call counterclockwise positive), and in Section 10.4 we will give it a full vector definition with the right-hand rule.

Checkpoint 10.4

A stubborn bolt will not budge, so you slip a length of pipe over the wrench handle to double its length, then push just as hard as before and still perpendicular to the handle.

a. By what factor does the torque on the bolt change?

b. Starting over with the original wrench, suppose instead you keep the handle length the same but push at 30° to the handle rather than perpendicular to it. By what factor does the torque change now?
c. Which of the two changes helps more? And why is a push 30° *past* perpendicular just as effective as one 30° short of it?

Answer

Torque as a vector

So far we have treated torque as a magnitude with a clockwise-or-counterclockwise sense. That is enough for motion in a plane, but torque is really a **vector**, and its full definition is the *cross product* of the position vector and the force, exactly the vector product introduced in Chapter 3:

Key Equation: Torque as a vector

$$\tau = \mathbf{r} \times \mathbf{F}, \qquad |\tau| = rF\sin\theta. \tag{10.13}$$

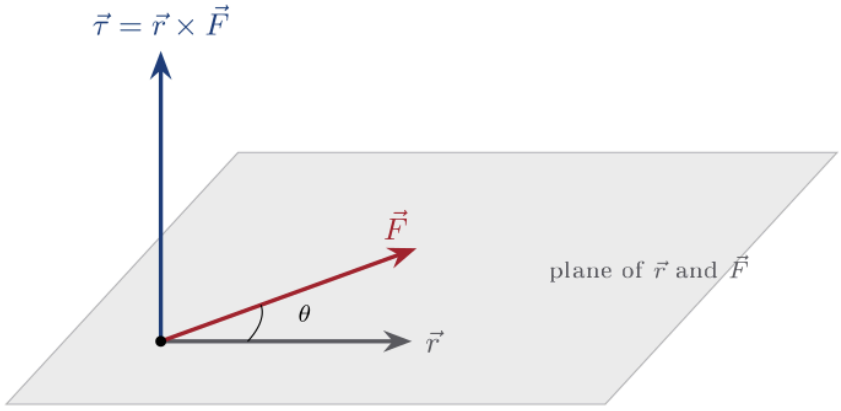


Figure 10.5. The torque $\tau = \mathbf{r} \times \mathbf{F}$ is perpendicular to the plane containing $\mathbf{r}$ and $\mathbf{F}$, with its direction given by the right-hand rule: point your fingers along $\mathbf{r}$, curl them toward $\mathbf{F}$, and your thumb gives τ. When $\mathbf{r}$ and $\mathbf{F}$ both lie in the page, τ points straight out of the page for a counterclockwise torque and into it for a clockwise one.

•Figure description

Torque as a vector cross product. Vectors r and F both lie in a shaded plane, drawn in perspective, with angle theta between them. The torque vector tau equals r cross F and points perpendicular to that plane, its direction given by the right-hand rule.

Here $\mathbf{r}$ runs from the axis to the point where the force is applied, and θ is the angle between $\mathbf{r}$ and $\mathbf{F}$ when the two are drawn *tail to tail*. The magnitude $rF\sin\theta$ is the lever-arm result we already know (equation (3.17)). The *direction* is given by the **right-hand rule**: point the fingers of your right hand along $\mathbf{r}$, curl them toward $\mathbf{F}$, and your thumb points along τ, perpendicular to the plane of $\mathbf{r}$ and $\mathbf{F}$ (Figure 10.5). For rotation in the xy-plane the torque points along $\pm z$: out of the page for a counterclockwise torque, into the page for a clockwise one. This is the precise meaning behind the counterclockwise-positive sign convention we have been using.

Note: Two ways to go wrong with the cross product

First, *order matters*. The cross product is anticommutative, so $\mathbf{F} \times \mathbf{r} = -\mathbf{r} \times \mathbf{F}$: writing the factors in the wrong order reverses the torque. The position vector always comes first. Second, θ is the angle between $\mathbf{r}$ and $\mathbf{F}$ *drawn tail to tail*, not the angle you happen to see in the sketch between the force and some nearby edge or wall. If you push on a wrench handle at 30° past perpendicular, then $\theta = 120^\circ$, not 30°.

The same right-hand rule promotes angular velocity to a vector. Curl the fingers of your right hand in the direction the body is turning; your thumb then points along ω, parallel to the rotation axis. A wheel spinning counterclockwise in the page has ω out of the page. With that convention, τ and ω point the same way when a body is speeding up and opposite ways when it is slowing down: the rotational echo of $\mathbf{F}$ and $\mathbf{v}$ for straight-line motion.

Example 10.3: Torque from the cross product

A force $\mathbf{F} = (12\ \hat{\imath} + 3\ \hat{\jmath})\ \mathrm{N}$ is applied at a point located at $\mathbf{r} = (1.4\ \hat{\imath} + 0.50\ \hat{\jmath})\ \mathrm{m}$, on an object pivoted about the z-axis. Find the torque about the axis.

Solution

Both vectors lie in the xy-plane ($r_z = F_z = 0$), so in the determinant of equation (3.19) every term that multiplies $\hat{\imath}$ or $\hat{\jmath}$ contains a z-component and vanishes. Only the z-component survives, $\tau_z = r_x F_y - r_y F_x$: the x of the first vector times the y of the second, *minus* the y of the first times the x of the second.

$$\tau = \mathbf{r} \times \mathbf{F} = \big[(1.4)(3) - (0.50)(12)\big]\hat{k}\,\mathrm{N \cdot m} = (4.2 - 6.0)\hat{k}\,\mathrm{N \cdot m} = \boxed{-1.8\,\hat{k}\ \mathrm{N \cdot m}}. \tag{10.4.1}$$

The order matters: $\mathbf{r}$ comes first because torque is defined as $\mathbf{r} \times \mathbf{F}$; reversing it would flip the sign.

The torque has magnitude $1.8\ \mathrm{N \cdot m}$ and points in the $-z$ direction (into the page), so it would tend to turn the object clockwise. Note that the two terms compete: the y-component of the force, acting at lever arm $r_x = 1.4\ \mathrm{m}$, tries to turn the object counterclockwise ($+4.2\ \mathrm{N \cdot m}$), while the x-component, acting at lever arm $r_y = 0.50\ \mathrm{m}$, tries to turn it clockwise ($-6.0\ \mathrm{N \cdot m}$). Here the second wins.

Checkpoint 10.5

A wheel lies flat in the plane of this page and is spinning clockwise.

a. In which direction does ω point?
b. A torque is applied that makes the wheel spin faster. In which direction does τ point?
c. Now the wheel slows down instead. Which of your two answers changes, and which does not?

Answer

Newton's second law for rotation

Torque is to rotation what force is to translation, and it enters an exactly analogous law. The reason is short enough to give here. Apply Newton's second law to one particle of the body in the tangential direction, $F_{t,i} = m_i a_{t,i}$, and multiply both sides by that particle's distance r_i from the axis. The left side becomes the torque on the particle, and on the right $a_{t,i} = r_i \alpha$ turns the expression into

$$\tau_i = r_i F_{t,i} = m_i r_i^2\, \alpha. \tag{10.4.2}$$

Every particle of a rigid body shares the same α, so summing over the whole body factors it out, and what is left of the sum is precisely the moment of inertia. The result is **Newton's second law for rotation**:

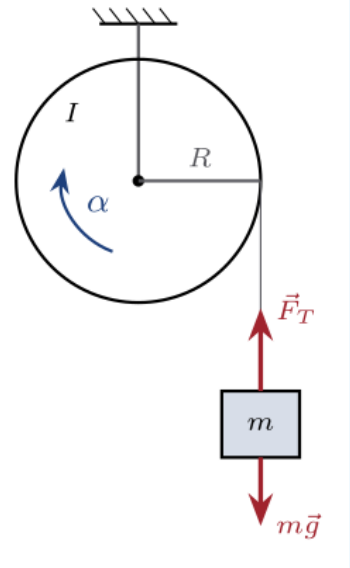
Key Equation: Newton's second law for rotation

$$\sum \tau = I\alpha. \tag{10.14}$$

Compare $\sum F = ma$: torque replaces force, moment of inertia replaces mass, and angular acceleration replaces linear acceleration. One condition is built into the derivation and easy to overlook: the torques on the left and the moment of inertia on the right must be taken about the *same* fixed axis.

Example 10.4: A hanging mass turning a wheel

A wheel of radius R and moment of inertia I turns on a frictionless horizontal axle. A light cord is wrapped around its rim, and from the free end hangs an object of mass m. When released, the object falls and the cord unwinds, spinning the wheel. Find the angular acceleration of the wheel, the linear acceleration of the object, and the cord tension.



•Figure description
An object of mass m hanging from a cord wrapped around a wheel of radius R and moment of inertia I. The cord tension acts on the object along with its weight, and the wheel has angular acceleration alpha.

Solution

Step 1: One equation for each object. Treat the two objects separately. For the falling block (taking down as positive), the weight points down and the tension up:

$$mg - F_T = ma. \tag{10.4.3}$$

For the wheel, the cord's tension acts at the rim with lever arm R, producing a torque $F_T R$, and Newton's second law for rotation gives

$$F_T R = I\alpha. \tag{10.4.4}$$

Step 2: The link between them. That is two equations in three unknowns (a, α, F_T), so we need one more relation. The cord does not slip, so the block's acceleration matches the rim's tangential acceleration: $a = R\alpha$, i.e. $\alpha = a/R$. Substituting into the wheel equation and dividing both sides by R:

$$F_T R = I\,\frac{a}{R} \implies F_T = \frac{Ia}{R^2}. \tag{10.4.5}$$

Step 3: Solve for a. Substitute this F_T into the block equation. Then gather both terms containing a on one side and factor it out:

$$mg - \frac{Ia}{R^2} = ma \implies mg = ma + \frac{Ia}{R^2} = a\left(m + \frac{I}{R^2}\right). \tag{10.4.6}$$

Divide both sides by the parenthesis, then multiply numerator and denominator by R^2 to clear the fraction-within-a-fraction:

$$a = \frac{mg}{m + I/R^2} = \boxed{\frac{mgR^2}{I + mR^2}}. \tag{10.4.7}$$

Step 4: The other two. Each follows by substituting this a:

$$\alpha = \frac{a}{R} = \boxed{\frac{mgR}{I + mR^2}}, \qquad F_T = \frac{Ia}{R^2} = \frac{I}{R^2}\cdot\frac{mgR^2}{I + mR^2} = \boxed{\frac{mgI}{I + mR^2}}. \tag{10.4.8}$$

In the tension, the R^2 in the numerator of a cancels the R^2 in the denominator of I/R^2.

Checks. If $I \to 0$ the wheel offers no resistance and $a \to mgR^2/(mR^2) = g$, $F_T \to 0$: the block is in free fall. A massive wheel (large I) makes a small. The units work too: I and mR^2 are both $\mathrm{kg \cdot m^2}$, so they can be added, and a comes out in the units of g.

Notice also that $F_T < mg$ for any finite I, since $I/(I + mR^2) < 1$. It has to be: the block is accelerating downward, so the upward pull of the cord cannot balance its weight. Setting $F_T = mg$ is the most common mistake in problems of this kind, and it would leave the block hanging motionless.

Rolling down a ramp

With $\sum \tau = I\alpha$ in hand we can return to the rolling wheel of Section 10.2. Set a wheel on an incline and it accelerates down the slope, spinning faster as it goes. Something must be exerting the torque that spins it up, and gravity cannot: gravity acts at the center of mass, so its lever arm about the center is zero. The normal force acts along a line through the center too. The only remaining candidate is **static friction** at the contact point, and it is what makes rolling on a ramp possible at all. On a perfectly frictionless ramp a wheel would slide down without ever beginning to turn.

Take a body of mass m, radius R, and moment of inertia written in the form $I_{\mathrm{cm}} = \beta m R^2$, where β is a pure number fixed by the shape: $\beta = 1$ for a hoop, $\frac{1}{2}$ for a solid disk or cylinder, $\frac{2}{5}$ for a solid sphere (Table 10.1). On an incline of angle θ, Newton's second law along the slope and about the center of mass gives two equations,

$$mg\sin\theta - F_f = ma_{\mathrm{cm}}, \qquad F_f R = I_{\mathrm{cm}}\alpha = \beta m R^2 \,\frac{a_{\mathrm{cm}}}{R}, \tag{10.4.9}$$

where the rolling condition $a_{\mathrm{cm}} = R\alpha$ has already been used in the second. The second equation says $F_f = \beta m a_{\mathrm{cm}}$; substituting into the first and solving,

Key Equation: Rolling down an incline

$$a_{\mathrm{cm}} = \frac{g\sin\theta}{1+\beta}, \qquad F_f = \frac{\beta}{1+\beta}\,mg\sin\theta, \qquad \mu_s \geq \frac{\beta\tan\theta}{1+\beta}. \tag{10.15}$$

The last condition comes from requiring $F_f \leq \mu_s F_N$ with $F_N = mg\cos\theta$: if the surface cannot supply that much friction, the body slips and the rolling condition fails. Notice what is *absent* from all three results. Neither m nor R appears, only the shape factor β. Two spheres of wildly different size and mass roll down the same ramp side by side, and a hoop ($\beta = 1$) always loses to a solid sphere ($\beta = \frac{2}{5}$). More of the hoop's mass sits far from its center, so friction must take a larger share of the downhill pull just to spin it up, leaving less to speed it down the slope. Section 11.1 tells the same story in terms of energy.

Checkpoint 10.6

A solid sphere ($\beta = \frac{2}{5}$) and a thin spherical shell ($\beta = \frac{2}{3}$) have the same mass and radius and are placed on the same ramp. Using equation (10.15) and without computing anything numerically, decide which one requires the larger coefficient of static friction to roll without slipping, and explain in words why the answer had to come out that way. What happens to both if the ramp is made steeper and steeper?

Answer

Example 10.5: An unfurling roll of paper

A roll of paper is held by its free end and released, unfurling as it falls (Figure 10.6). The roll has outer radius $R = 6.0\ \mathrm{cm}$, inner radius $r = 1.8\ \mathrm{cm}$, and mass $m = 200\ \mathrm{g} = 0.200\ \mathrm{kg}$, and it falls $s = 3.0\ \mathrm{m}$. Treating it as a uniform thick-walled (annular) cylinder, for which $I = \frac{1}{2}m(R^2 + r^2)$, and assuming R and m stay essentially constant, find (a) its translational acceleration, (b) its angular acceleration, (c) its final translational speed, (d) its final angular speed, and (e) the tension in the sheet. Take $g = 9.8\ \mathrm{m/s^2}$.

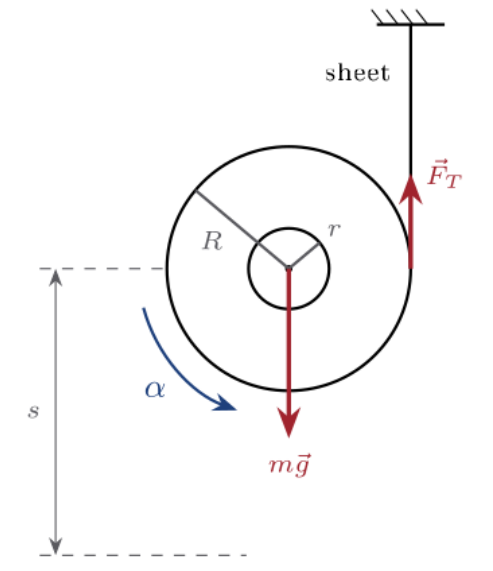


Figure 10.6. A roll held by its free end unfurls as it falls. The sheet leaves the outer edge, so its tension acts at radius R and the unwinding constraint is $a = R\alpha$.

•Figure description

A roll of paper with outer radius R and inner radius r hangs from a sheet fixed above it; the sheet leaves the roll at its right edge. The tension F sub T acts upward at the rim where the sheet leaves, the weight m g acts downward at the center, and a curved arrow shows the counterclockwise angular acceleration alpha. A dimension line marks the distance s the roll falls.

Solution

Step 1: The physical picture. The sheet plays exactly the part the ground plays for a rolling wheel: the point where the paper leaves the roll is instantaneously at rest, so the roll "rolls" down the sheet. The tension therefore acts at radius R, and the unwinding constraint is $a = R\alpha$.

Step 2: Two equations. Newton's second law for the falling center of mass (down positive), and for rotation about the center:

$$mg - F_T = ma, \qquad F_T R = I\alpha = I\,\frac{a}{R}. \tag{10.4.10}$$

Dividing the second equation by R gives $F_T = Ia/R^2$, the same structure as the hanging mass and wheel earlier in this section.

Step 3: Solve for a. Substitute this F_T into the first equation, gather the a terms, and factor:

$$mg - \frac{Ia}{R^2} = ma \implies mg = a\left(m + \frac{I}{R^2}\right). \tag{10.4.11}$$

Now divide both sides by m (so that $m/m = 1$ and $I/R^2 \div m = I/(mR^2)$), and then by the parenthesis:

$$a = \frac{g}{1 + I/(mR^2)}. \tag{10.4.12}$$

Writing it this way isolates the dimensionless ratio $I/(mR^2)$, which depends only on the *shape* of the roll. With $I = \frac{1}{2}m(R^2 + r^2)$, the m cancels, and dividing each term in the parenthesis by R^2 gives

$$\frac{I}{mR^2} = \frac{\frac{1}{2}m(R^2 + r^2)}{mR^2} = \frac{1}{2}\left(1 + \frac{r^2}{R^2}\right) = \frac{1}{2}\left(1 + \left(\frac{1.8}{6.0}\right)^2\right) = \frac{1}{2}(1 + 0.09) = 0.545. \tag{10.4.13}$$

The two radii appear only as a ratio, so there is no need to convert centimeters to meters for this step. Then:

(a) Translational acceleration.

$$a = \frac{9.8\,\mathrm{m/s^2}}{1 + 0.545} = \frac{9.8}{1.545}\,\mathrm{m/s^2} \approx \boxed{6.3\,\mathrm{m/s^2}}. \tag{10.4.14}$$

(b) Angular acceleration. From $a = R\alpha$, divide by R, which now *must* be in meters ($R = 0.060\ \text{m}$):

$$\alpha = \frac{a}{R} = \frac{6.34\,\text{m/s}^2}{0.060\,\text{m}} \approx \boxed{106\,\text{rad/s}^2}. \tag{10.4.15}$$

(c) Final translational speed. The acceleration is constant and the roll starts from rest, so $v^2 = 0 + 2as$:

$$v = \sqrt{2(6.34\,\text{m/s}^2)(3.0\,\text{m})} = \sqrt{38.0\,\text{m}^2/\text{s}^2} \approx \boxed{6.2\,\text{m/s}}. \tag{10.4.16}$$

(d) Final angular speed. The unwinding constraint holds for speeds too, $v = R\omega$:

$$\omega = \frac{v}{R} = \frac{6.17\,\text{m/s}}{0.060\,\text{m}} \approx \boxed{103\,\text{rad/s}}. \tag{10.4.17}$$

(e) Tension in the sheet. From the first equation of Step 2:

$$F_T = mg - ma = m(g - a) = (0.200\,\text{kg})(9.8 - 6.34)\,\text{m/s}^2 \approx \boxed{0.69\,\text{N}}. \tag{10.4.18}$$

Checks. The tension came out well below the roll's weight of $1.96\ \text{N}$, as it must for a roll that is accelerating downward. Part (e) can also be checked from the rotation equation: $F_T = Ia/R^2 = m(0.545)a = (0.200)(0.545)(6.34) = 0.69\ \text{N}$. And the moment of inertia behaves sensibly at the extremes: as $r \to 0$ the formula gives $\frac{1}{2}mR^2$, the solid cylinder, while as $r \to R$ it gives mR^2, the thin hoop.

Note: Taking torques about an accelerating center of mass

The example above applied $\sum \tau = I_{\text{cm}}\alpha$ about the center of the roll: a point that is itself accelerating downward, and therefore not the origin of an inertial frame. That step is nonetheless legitimate, and it is worth knowing why: for *any* rigid body, torques and moments of inertia taken about an axis through the center of mass obey $\sum \tau_{\text{cm}} = I_{\text{cm}}\alpha$ even when the center of mass accelerates. The center of mass is the one accelerating point for which this works. About any other accelerating point, extra terms appear and the simple law fails.

Answers to Checkpoints

Checkpoint 10.4. (a) Torque is $rF\sin\theta$. Doubling r with the same perpendicular push ($\sin 90^\circ = 1$) doubles the torque, a factor of 2. (b) With the original handle, pushing at 30° to it multiplies the torque by $\sin 30^\circ = 0.5$, only half the torque of a perpendicular push. (c) Lengthening the handle helps more: it doubles the torque, while the angled push halves it. A push 30° past perpendicular makes 120° with the handle, and one 30° short makes 60°; since $\sin 120^\circ = \sin 60^\circ = 0.866$, both give the same torque. Only the component of the force perpendicular to the handle turns the bolt, and it is the same in both cases.

Checkpoint 10.5. (a) Curl the fingers of your right hand in the clockwise sense of rotation: your thumb points *into* the page, so ω points into the page. (b) Speeding up means ω grows in the direction it already points, so τ (like α) points into the page too. (c) If the wheel slows, τ reverses and points out of the page, while ω still points into the page. The direction of ω is set by the sense of rotation; the direction of τ is set by whether the rotation is speeding up or slowing down.

Checkpoint 10.6. The shell. The required coefficient is $\mu_s \geq \beta\tan\theta/(1+\beta)$, and the factor $\beta/(1+\beta)$ increases with β: it is $2/7 \approx 0.29$ for the solid sphere and $2/5 = 0.40$ for the shell. In words, the shell carries its mass farther from its center, so a larger torque, and therefore a larger friction force, is needed to spin it up in step with its acceleration down the ramp. As the ramp steepens, $\tan\theta$ grows without limit, so eventually both must slip; the shell reaches that point first. (At $\theta = 90^\circ$ there is no normal force at all, so no friction is possible.)

This page titled 10.4: Torque and the Right-Hand Rule is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.S: Rotation I (Summary)

Summary

- Rotation is described by **angular displacement** θ, **angular velocity** ω, and **angular acceleration** α. Linear and angular quantities are linked by $s = r\theta$, $v = r\omega$, $a_t = r\alpha$, and $a_c = r\omega^2$, every one of which requires the angle in radians.
- For constant α, the **rotational kinematic equations** match the linear ones with angular symbols substituted.
- In **rolling without slipping**, $v_{\text{cm}} = R\omega$ and $a_{\text{cm}} = R\alpha$; the contact point is instantaneously at rest and the top of the body moves at $2v_{\text{cm}}$.
- **Moment of inertia** $I = \sum m_i r_i^2 = \int r^2 \, dm$ measures resistance to angular acceleration; it depends on how mass is distributed about the axis, not just on the amount of mass.
- The **parallel-axis theorem** $I = I_{\text{cm}} + Md^2$ gives I about any axis parallel to one through the center of mass; I is always smallest about the center of mass itself.
- **Torque** $\tau = rF\sin\theta = r_\perp F = rF_\perp$ is the rotational effect of a force: only the perpendicular part of the force does any turning. As a vector, $\tau = \mathbf{r} \times \mathbf{F}$, with direction set by the right-hand rule (out of the page for counterclockwise). Angular velocity ω is a vector by the same rule.
- **Newton's second law for rotation** is $\sum \tau = I\alpha$, with both sides referred to the same fixed axis, or to the center of mass even when it accelerates.
- **Static friction** supplies the torque that spins a body up on a ramp. Writing $I_{\text{cm}} = \beta mR^2$, a body rolling down an incline has $a_{\text{cm}} = g\sin\theta/(1+\beta)$ and requires $\mu_s \geq \beta\tan\theta/(1+\beta)$: results that depend on the shape factor β but on neither the mass nor the radius.

Key Equations

Relating linear and angular motion

$$s = r\theta, \qquad v = r\omega, \qquad a_t = r\alpha, \qquad a_c = r\omega^2 = \frac{v^2}{r} \tag{10.S.1}$$

Rotational kinematic equations (constant α)

$$\begin{aligned} \omega &= \omega_0 + \alpha\, t, \\ \Delta\theta &= \omega_0\, t + \tfrac{1}{2}\alpha\, t^2, \\ \omega^2 &= \omega_0^2 + 2\alpha\, \Delta\theta, \\ \Delta\theta &= \tfrac{1}{2}(\omega + \omega_0)\, t \end{aligned} \tag{10.S.2}$$

Rolling without slipping

$$x_{\text{cm}} = R\theta, \qquad v_{\text{cm}} = R\omega, \qquad a_{\text{cm}} = R\alpha \tag{10.S.3}$$

Moment of inertia

$$I = \sum_i m_i r_i^2 \quad \text{(particles)}, \qquad I = \int r^2\, dm \quad \text{(continuous body)} \tag{10.S.4}$$

Parallel-axis theorem

$$I = I_{\text{cm}} + Md^2 \tag{10.S.5}$$

Torque

$$\tau = rF\sin\theta = r_{\perp}F \tag{10.S.6}$$

Torque as a vector

$$\tau = \mathbf{r}\times\mathbf{F}, \qquad |\tau| = rF\sin\theta \tag{10.S.7}$$

Newton's second law for rotation

$$\sum\tau = I\alpha \tag{10.S.8}$$

Rolling down an incline

$$a_{\text{cm}} = \frac{g\sin\theta}{1+\beta}, \qquad F_f = \frac{\beta}{1+\beta}\,mg\sin\theta, \qquad \mu_s \geq \frac{\beta\tan\theta}{1+\beta} \tag{10.S.9}$$

This page titled 10.S: Rotation I (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

10.E: Rotation I (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 10.1 — Rotational Kinematics

10.1 ★ A wheel turns through $12\ \text{rad}$ in $3.0\ \text{s}$ at constant angular velocity.

a. What is its angular velocity in rad/s?
b. How many revolutions is $12\ \text{rad}$?

10.2 ★ *[Conceptual.]* A merry-go-round turns at a steady rate. Compare, for a child near the center and a child at the outer edge: (a) their angular velocities, and (b) their linear (tangential) speeds. Explain using $v = r\omega$.

10.3 ★★ A centrifuge accelerates from rest to $12000\ \text{rpm}$ in $4.0\ \text{s}$ at constant angular acceleration.

a. Find the angular acceleration in rad/s^2.
b. Through how many revolutions does it turn during this time?

Section 10.2 — Rolling Motion

10.4 ★ A wheel of radius $0.35\ \text{m}$ rolls without slipping at a center-of-mass speed of $8.0\ \text{m/s}$.

a. What is its angular speed?
b. What is the speed of the point at the top of the wheel?

10.5 ★★ *[Real-world: transportation.]* A car's tires have radius $0.33\ \text{m}$. The car accelerates from rest to $27\ \text{m/s}$ in $9.0\ \text{s}$.

a. Find the final angular velocity of a tire.
b. Find the angular acceleration of a tire.

Section 10.3 — Rotational Inertia

10.6 ★ Two $0.50\ \text{kg}$ masses are attached to the ends of a light rod of length $1.2\ \text{m}$. Find the moment of inertia of this dumbbell about an axis through its center, perpendicular to the rod.

10.7 ★★ *[Conceptual.]* A solid disk and a thin hoop have the same mass and radius.

a. Which has the larger moment of inertia about its central axis? Why?
b. If the same torque is applied to each, which gains angular speed faster?

10.8 ★★ Four $2.0\ \text{kg}$ point masses sit at the corners of a square of side $1.0\ \text{m}$. Find the moment of inertia about an axis perpendicular to the square through its center. *(Each corner is the same distance from the center.)*

10.9 ★★★ Using $I = \int r^2\ dm$, show that the moment of inertia of a uniform thin hoop of mass M and radius R about its central axis is MR^2, and explain why this is the largest possible value of I for any body of that mass and radius.

Section 10.4 — Torque and the Right-Hand Rule

10.10 ★ A mechanic pushes with $180\ \text{N}$ at the end of a wrench $0.25\ \text{m}$ long, perpendicular to the handle. What torque is applied to the bolt?

10.11 ★ *[Conceptual.]* Explain why a door is much easier to open when you push near the outer edge than near the hinges,

and why pushing straight toward the hinges does nothing at all. Use the idea of the lever arm.

10.12 ★★ A light rod $0.60\ \text{m}$ long is pivoted at one end. A $40\ \text{N}$ force pulls on the free end at an angle of 30° to the rod.

a. Find $F_\perp$, the component of the force perpendicular to the rod, and use it to find the torque about the pivot.
b. Now keep the whole force and find $r_\perp$, the perpendicular distance from the pivot to the line along which the force acts. Use it to find the torque again. Do the two methods agree?
c. The force is turned so that it pulls straight along the rod, away from the pivot. What are $F_\perp$, $r_\perp$ and the torque now?

10.13 ★★★ A $15\ \text{N}$ force is applied at the end of a $0.40\ \text{m}$ wrench. Find the torque magnitude if the force is applied (a) perpendicular to the wrench, (b) at 60° to the wrench, and (c) along the wrench. Comment on the trend.

10.14 ★ A position vector $\mathbf{r} = (3.0\ \hat{\imath})\ \text{m}$ locates the point where a force $\mathbf{F} = (4.0\ \hat{\jmath})\ \text{N}$ is applied. Find the torque $\tau = \mathbf{r} \times \mathbf{F}$, including its direction.

10.15 ★ *[Conceptual.]* You loosen a bolt by pushing down on the end of a horizontal wrench. Use the right-hand rule to determine the direction of the torque vector, and explain how it relates to which way the bolt turns.

10.16 ★★ A force $\mathbf{F} = (2.0\ \hat{\imath} + 5.0\ \hat{\jmath})\ \text{N}$ acts at the point $\mathbf{r} = (4.0\ \hat{\imath} - 1.0\ \hat{\jmath})\ \text{m}$. Find the torque about the origin (magnitude and direction).

10.17 ★★★ A grindstone is a solid disk of mass $8.0\ \text{kg}$ and radius $0.20\ \text{m}$. A constant torque gives it an angular acceleration of $6.0\ \text{rad/s}^2$.

a. Find the moment of inertia of the grindstone.
b. Find the applied torque.
c. What tangential force at the rim would produce this torque?

10.18 ★★★ *[Real-world: engineering.]* A bucket of mass $3.0\ \text{kg}$ hangs from a light rope wound around a solid cylindrical drum of mass $5.0\ \text{kg}$ and radius $0.15\ \text{m}$ on a frictionless axle. The bucket is released from rest. Find the bucket's acceleration and the rope tension. *(Use $I = \frac{1}{2}MR^2$ for the drum.)*

Challenge Problems

10.19 ★★★ A uniform solid disk of mass M and radius R can be built up out of thin concentric rings. A ring of radius r and width dr has area $2\pi r\ dr$, and every point on it lies the same distance r from the central axis.

a. Write the mass dm of such a ring in terms of M, R, r, and dr.
b. Use this with $I = \int r^2\ dm$ to show that the disk's moment of inertia about its central axis is $\frac{1}{2}MR^2$, as claimed in Table 10.1.
c. Now suppose the disk is *not* uniform: its area density grows linearly with radius, $\sigma(r) = \sigma_0 r/R$. Show that its moment of inertia about the same axis is $\frac{3}{5}MR^2$, and explain in one sentence why this had to come out larger than $\frac{1}{2}MR^2$.

10.20 ★★★ A solid cylinder ($I = \frac{1}{2}MR^2$) of mass $4.0\ \text{kg}$ and radius $0.10\ \text{m}$ rolls without slipping down an incline of angle 30°.

a. Using Newton's second law for translation and rotation together, find the acceleration of its center of mass.
b. Find the friction force acting on the cylinder.
c. What is the minimum coefficient of static friction needed to prevent slipping?

This page titled 10.E: Rotation I (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11: ROTATION II

Chapter 10 described rotation, measured how strongly a body resists being turned, and gave rotation its own version of Newton's second law. This chapter completes the parallel with translation. We first find the **kinetic energy** of a spinning body, including the combined translation and rotation of a rolling wheel, and the **work** and power delivered by a torque. Then we meet **angular momentum** (the rotational partner of linear momentum) and its conservation law, one of the deepest principles in physics, which governs everything from a spinning skater to the collapse of a star into a pulsar. We close with the slow precession of a spinning gyroscope.

Learning Objectives

- Calculate rotational kinetic energy, and the total kinetic energy of a body rolling without slipping.
- Use energy conservation to find the speed of a body rolling down a ramp.
- Calculate the work done and the power delivered by a torque, and apply the rotational work–kinetic energy theorem.
- Define angular momentum for a particle and for a rigid body, and relate net torque to its rate of change.
- Apply conservation of angular momentum to systems with no external torque.
- Describe the precession of a spinning gyroscope.

11.1: Rotational Kinetic Energy

11.2: Rotational Work

11.3: Angular Momentum

11.4: Conservation of Angular Momentum

11.5: Precession

11.S: Rotation II (Summary)

11.E: Rotation II (Problems)

This page titled 11: Rotation II is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.1: Rotational Kinetic Energy

A spinning body has kinetic energy even if it is going nowhere, because all of its particles are moving. Each particle has its own speed v_i, but all of them are built from the single ω that the whole body shares, through $v_i = r_i\omega$. Adding up $\frac{1}{2}m_i v_i^2$ over every particle and substituting,

$$K = \sum_i \tfrac{1}{2} m_i v_i^2 = \sum_i \tfrac{1}{2} m_i (r_i\omega)^2 = \tfrac{1}{2}\Big(\sum_i m_i r_i^2\Big)\omega^2. \tag{11.1.1}$$

The sum in parentheses is just the moment of inertia, so the **rotational kinetic energy** is

Key Equation: Rotational kinetic energy

$$K_{\text{rot}} = \tfrac{1}{2} I\omega^2. \tag{11.1}$$

This is the perfect analog of $K = \frac{1}{2}mv^2$, with I in place of m and ω in place of v. It is measured in joules and counts as part of a system's mechanical energy, so conservation of energy (Chapter 8) applies to rotating bodies too.

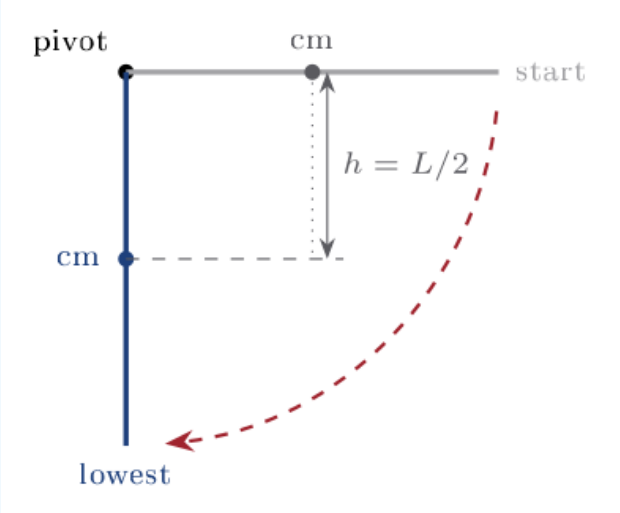
Checkpoint 11.1

A thin hoop and a solid disk have the same mass and the same radius, and at this moment they carry the same rotational kinetic energy. Which one is turning faster, and by what factor? *(Their moments of inertia are MR^2 and $\frac{1}{2}MR^2$.)*

Answer

Example 11.1: A falling rod

A uniform rod of length L and mass M pivots without friction about one end. It is released from rest in the horizontal position. (a) Find its angular speed when it reaches the vertical (lowest) position. (b) Find the tangential speed of the center of mass and of the lowest tip at that moment.



•Figure description
A rod pivoted at one end, shown released from a horizontal starting position and swinging down to its lowest vertical position, with a dashed arc showing the path of its free end. The center of mass, at the middle of the rod, falls a height h equal to half the rod's length.

Solution

(a) Use energy conservation; the frictionless pivot does no work. As the rod swings down, its center of mass (at the middle) drops a height $h = L/2$, so the potential energy lost is $Mg(L/2)$. (Even though different parts of the rod drop different amounts, the potential energy of the whole rod is Mg times the height of its center of mass.) This becomes rotational kinetic energy, with the rod pivoting about its end, $I = \frac{1}{3}ML^2$:

$$Mg\frac{L}{2} = \tfrac{1}{2}I\omega^2 = \tfrac{1}{2}\left(\tfrac{1}{3}ML^2\right)\omega^2 = \tfrac{1}{6}ML^2\omega^2. \tag{11.1.2}$$

To isolate ω^2, multiply both sides by 6 and divide both sides by ML^2. The mass cancels, and one power of L cancels:

$$\omega^2 = \frac{6\,Mg\,(L/2)}{ML^2} = \frac{3MgL}{ML^2} = \frac{3g}{L} \implies \omega = \boxed{\sqrt{\frac{3g}{L}}}. \tag{11.1.3}$$

The units check: $(\ \mathrm{m/s^2})/\ \mathrm{m} = \ \mathrm{s^{-2}}$, whose square root is $\ \mathrm{s^{-1}}$, the unit of angular speed.

(b) The tangential speed is $v = r\omega$. The center of mass is at $r = L/2$ and the tip at $r = L$. To combine a factor outside the root with the root, bring it inside as its square, $L = \sqrt{L^2}$:

$$v_{\mathrm{cm}} = \frac{L}{2}\sqrt{\frac{3g}{L}} = \frac{1}{2}\sqrt{\frac{3gL^2}{L}} = \boxed{\tfrac{1}{2}\sqrt{3gL}}, \qquad v_{\mathrm{tip}} = L\sqrt{\frac{3g}{L}} = \boxed{\sqrt{3gL}}. \tag{11.1.4}$$

The tip moves exactly twice as fast as the center of mass, since it is twice as far from the axis.

Check. A ball dropped from rest through the tip's fall of L would reach $\sqrt{2gL}$. The tip of the rod goes faster, $\sqrt{3gL}$, because the rest of the rod, falling less far, pushes it along. Surprising, but a well-known result: a falling chimney or a toppling meter stick outruns free fall at its far end.

Kinetic energy of a rolling body

A body rolling without slipping (Section 10.2) both translates and rotates, so it carries both kinds of kinetic energy. Its total is the sum,

Key Equation: Kinetic energy of a rolling body

$$K = \tfrac{1}{2}mv_{\mathrm{cm}}^2 + \tfrac{1}{2}I_{\mathrm{cm}}\omega^2, \tag{11.2}$$

with the two terms linked by equation (10.9). This is why a ball rolling down a ramp arrives with a smaller center-of-mass speed than one that slides down a frictionless ramp: some of the released potential energy must go into spin.

For a body rolling down a ramp, energy conservation gives the speed at the bottom in one line. Write the moment of inertia as $I_{\mathrm{cm}} = \beta mR^2$, as in Section 10.4. Falling a height h releases mgh, and by equations (10.9) and (11.2),

$$mgh = \tfrac{1}{2}mv_{\mathrm{cm}}^2 + \tfrac{1}{2}\left(\beta mR^2\right)\left(\frac{v_{\mathrm{cm}}}{R}\right)^2 = \tfrac{1}{2}mv_{\mathrm{cm}}^2(1+\beta) \qquad \implies \qquad v_{\mathrm{cm}} = \sqrt{\frac{2gh}{1+\beta}}. \tag{11.1.5}$$

This agrees with the force-and-torque result of equation (10.15): an acceleration $g\sin\theta/(1+\beta)$ along a slope of length $h/\sin\theta$ gives $v_{\mathrm{cm}}^2 = 2gh/(1+\beta)$. The energy picture explains why a hoop ($\beta = 1$) loses the race to a solid sphere ($\beta = \frac{2}{5}$): the hoop must divert a larger share of the released energy into spin.

Note: Static friction does no work in rolling

Using energy conservation on a rolling body may look illegal, since a friction force is plainly acting. It is not, and the reason is the one fact that defines rolling: the contact point is instantaneously *at rest*. A force does work only when its point of application moves, so the static friction at the contact patch does no work and dissipates no energy: it merely transfers energy between the translational and rotational accounts. *Kinetic* friction, which acts when a body skids, is a different matter entirely: there the surfaces slide past one another and mechanical energy is lost as heat.

Answers to Checkpoints

Checkpoint 11.1. The disk. Equal kinetic energies mean $\frac{1}{2}I_{\text{hoop}}\omega_{\text{hoop}}^2 = \frac{1}{2}I_{\text{disk}}\omega_{\text{disk}}^2$, so $\omega_{\text{disk}}^2/\omega_{\text{hoop}}^2 = I_{\text{hoop}}/I_{\text{disk}} = MR^2/(\frac{1}{2}MR^2) = 2$. Taking the square root, the disk turns faster by a factor of $\sqrt{2} \approx 1.41$. It needs the higher angular speed because its mass is concentrated nearer the axis, where it moves more slowly.

This page titled 11.1: Rotational Kinetic Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.2: Rotational Work

To speed up a rotating body, to increase its rotational kinetic energy, a torque must do work on it, just as a force does work to change translational kinetic energy. As a torque turns a body through a small angle $d\theta$, it does work $\tau\ d\theta$, so in general $W = \int \tau\ d\theta$, and for a constant torque this is simply

Key Equation: Work done by a torque

$$W = \tau\,\Delta\theta, \qquad W_{\text{net}} = \Delta K_{\text{rot}} = \tfrac{1}{2}I\omega^2 - \tfrac{1}{2}I\omega_0^2. \tag{11.3}$$

The second statement is the **work–kinetic energy theorem** in rotational form: the net work done by *all* the torques equals the change in rotational kinetic energy, exactly mirroring $W_{\text{net}} = \Delta K$ from Chapter 8. Keep the two apart. The quantity $W = \tau\ \Delta\theta$ is the work done by a *single* torque; only when that torque is the only one acting is it also the net work.

Example 11.2: Spinning up a potter's wheel

A potter's wheel with moment of inertia $0.30\ \text{kg}\cdot\text{m}^2$ starts from rest. A constant torque of $18\ \text{N}\cdot\text{m}$ acts through an angular displacement of $5.0\ \text{rad}$. Find the work done and the wheel's final angular speed.

Solution

Work. For a constant torque acting through an angle, the rotational analog of $W = Fd$ is

$$W = \tau\,\Delta\theta = (18\,\text{N}\cdot\text{m})(5.0\,\text{rad}) = 90\,\text{J}. \tag{11.2.1}$$

(The radian is dimensionless, so $\text{N}\cdot\text{m}$ times radians is simply joules.)

Final speed. By the rotational work–kinetic energy theorem, with $\omega_i = 0$, all of this work becomes rotational kinetic energy. Multiply both sides by 2, divide by I, and take the square root:

$$W = \tfrac{1}{2}I\omega_f^2 \implies \omega_f^2 = \frac{2W}{I} \implies \omega_f = \sqrt{\frac{2W}{I}}, \tag{11.2.2}$$

so

$$\omega_f = \sqrt{\frac{2(90\,\text{J})}{0.30\,\text{kg}\cdot\text{m}^2}} = \sqrt{600\,\text{s}^{-2}} \approx \boxed{24\,\text{rad/s}}. \tag{11.2.3}$$

Check. Do it with rotational kinematics instead. The angular acceleration is $\alpha = \tau/I = 18/0.30 = 60\ \text{rad/s}^2$, and $\omega_f^2 = \omega_i^2 + 2\alpha\ \Delta\theta = 0 + 2(60)(5.0) = 600\ \text{rad}^2/\text{s}^2$: the same $\sqrt{600}$. This is exactly parallel to the linear example "Speed from net work" in Chapter 8.

Checkpoint 11.2

A motor drives a wheel with a constant torque while a friction torque half that size opposes the motion, and the wheel turns through an angle $\Delta\theta$.

a. What is the sign of the work done by each of the two torques?
b. Which quantity equals ΔK_{rot}: the motor's work, or the net work?
c. If the two torques were instead equal in size, what would the wheel's angular speed be doing?

Answer

A torque that keeps acting delivers energy at some rate, and the rotational partner of $P = Fv$ drops straight out of $W = \tau\ \Delta\theta$.

Divide the work by the time taken, and the angular displacement per unit time is simply the angular velocity:

Key Equation: Power delivered by a torque

$$P = \tau\omega. \tag{11.4}$$

With τ in newton-meters and ω in radians per second, P comes out in watts. This equation is why motors and engines are always rated by torque *and* speed together rather than by either alone: a given power can be delivered as a large torque at a low speed or a small torque at a high speed, which is exactly what a gearbox trades between.

The parallel between translation and rotation, which has run through this chapter and the last, is worth collecting in one place (Table 11.1). Every rotational law is its linear counterpart with the matching symbols swapped in.

Table 11.1. The correspondence between translational and rotational quantities.

Translation	Rotation
position x	angle θ
velocity v	angular velocity ω
acceleration a	angular acceleration α
mass m	moment of inertia I
force F	torque τ
$\sum F = ma$	$\sum \tau = I\alpha$
$K = \frac{1}{2}mv^2$	$K_{\text{rot}} = \frac{1}{2}I\omega^2$
$W = Fd$	$W = \tau\ \Delta\theta$
$P = Fv$	$P = \tau\omega$

Answers to Checkpoints

Checkpoint 11.2. (a) The motor's torque acts in the direction of rotation, so its work is positive; the friction torque opposes the rotation, so its work is negative. (b) The *net* work (the sum of the two, here half the motor's work) equals ΔK_{rot}, just as the net work on a particle equals its ΔK. (c) Equal and opposite torques do zero net work, so the rotational kinetic energy does not change: the wheel turns at constant angular speed.

This page titled 11.2: Rotational Work is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.3: Angular Momentum

Just as linear momentum $\mathbf{p} = m\mathbf{v}$ measures translational motion, **angular momentum** $\mathbf{L}$ measures rotational motion. For a single particle it is the cross product of position and momentum; for a rigid body rotating about a fixed axis it takes the same tidy form as everything else in rotation, with I in place of m and ω in place of v:

Key Equation: Angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad |\mathbf{L}| = rmv\sin\theta; \qquad L = I\omega \text{ (rigid body)}. \tag{11.5}$$

Angular momentum has units of $\text{kg} \cdot \text{m}^2/\text{s}$ (equivalently $\text{J} \cdot \text{s}$), and like torque it is a vector whose direction follows the right-hand rule; for a rigid body it lies along the rotation axis, parallel to ω.

The factor $\sin\theta$ in equation (11.5) can be grouped in the same two ways the torque could be, and for the same reason:

$$L = r_\perp mv = r\, mv_\perp, \tag{11.6}$$

where $r_\perp = r\sin\theta$ is the perpendicular distance from the chosen point to the particle's *line of motion*, and $v_\perp = v\sin\theta$ is the component of velocity perpendicular to $\mathbf{r}$. Read either way, only motion that carries the particle *around* the point contributes; motion straight toward or away from the point contributes nothing at all.

Note: Angular momentum depends on the point you choose

Linear momentum is a property of a particle alone. Angular momentum is not: it is defined relative to a point (or an axis), because $\mathbf{r}$ is measured from that point. The same particle, at the same instant, has different angular momenta about different origins, and about a point lying on its own line of motion, its angular momentum is exactly zero. Whenever you write down an angular momentum, name the point it is about, and then keep using that same point throughout the problem.

Torque and angular momentum are connected by the rotational version of Newton's second law. Just as $\mathbf{F} = d\mathbf{p}/dt$, the net torque equals the rate of change of angular momentum,

Key Equation: Newton's second law (angular momentum form)

$$\sum \tau = \frac{d\mathbf{L}}{dt}. \tag{11.7}$$

For a rigid body of fixed moment of inertia this reduces to $\sum\tau = I\ d\omega/dt = I\alpha$, recovering the law of Chapter 10. Equation (11.7) is the more general statement, though, and it still applies when I itself changes, which is what makes the next section possible.

Example 11.3: A particle traveling in a straight line

A $2.0\ \text{kg}$ particle moves at a constant $3.0\ \text{m/s}$ in the $+x$ direction along the line $y = 1.5\ \text{m}$. Find its angular momentum (a) about the origin at the moment it crosses $x = 0$, (b) about the origin later, when it has reached $x = 4.0\ \text{m}$, and (c) about the point $(0,\ 1.5\ \text{m})$.

Solution

The particle's momentum is $\mathbf{p} = m\mathbf{v} = (6.0\ \hat{\imath})\ \text{kg} \cdot \text{m/s}$ throughout.

(a) Here $\mathbf{r} = (1.5\ \hat{\jmath})\ \text{m}$, so $L_z = r_x p_y - r_y p_x = (0)(0) - (1.5)(6.0) = \boxed{-9.0\ \text{kg} \cdot \text{m}^2/\text{s}}$, i.e. $9.0\ \text{kg} \cdot \text{m}^2/\text{s}$ into the page. The same answer follows in one step from equation (11.6): the line of motion passes a perpendicular distance $r_\perp = 1.5\ \text{m}$ from the origin, so $L = (1.5)(2.0)(3.0) = 9.0\ \text{kg} \cdot \text{m}^2/\text{s}$.

(b) Now $\mathbf{r} = (4.0\ \hat{\imath} + 1.5\ \hat{\jmath})\ \text{m}$, and

$$L_z = (4.0)(0) - (1.5)(6.0) = \boxed{-9.0\ \text{kg} \cdot \text{m}^2/\text{s}}, \tag{11.3.1}$$

unchanged. It had to be: no force acts on the particle, so no torque acts about any point, and by equation (11.7) its angular momentum cannot change. The perpendicular distance $r_\perp$ from the origin to the line of motion is a fixed geometric property of that line, and it is $r_\perp$, not the ever-growing r, that sets L.

(c) About the point $(0, 1.5)$ the particle's line of motion passes directly through the reference point, so $r_\perp = 0$ and $\boxed{L = 0}$ at every instant.

That a particle traveling in a perfectly straight line should have any angular momentum at all can feel strange, but part (b) shows why the definition insists on it: angular momentum is conserved when no torque acts, and a free particle is the simplest case of all.

Checkpoint 11.3

A satellite travels in a circular orbit around the Earth; later, another travels in a highly elliptical one.

a. For each, is there a torque on the satellite about the center of the Earth? (Gravity always points straight at that center.)
b. Which of the two moves at constant speed? And is angular momentum about the Earth's center constant for one of them, for both, or for neither?
c. Use equation (11.6) to argue that the elliptical satellite must move fastest when it is closest to Earth.

Answer

Answers to Checkpoints

Checkpoint 11.3. (a) No, for either one. Gravity points along the line to the Earth's center, so $\mathbf{r}$ and $\mathbf{F}$ are antiparallel and $\tau = \mathbf{r} \times \mathbf{F} = 0$ about that point. (b) Only the circular orbit has constant speed; the elliptical satellite speeds up and slows down. Yet angular momentum about the Earth's center is constant for *both*, because neither experiences a torque about that point. (c) With $L = r_\perp mv$ constant, a smaller perpendicular distance must be matched by a larger speed. At the closest point the velocity is perpendicular to the line to the Earth, so $r_\perp$ equals the full distance, which is then at its minimum: the speed must be at its maximum there. This is Kepler's second law.

This page titled 11.3: Angular Momentum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.4: Conservation of Angular Momentum

Equation (11.7) delivers a conservation law as powerful as conservation of linear momentum. If the net external torque on a system is zero, its angular momentum cannot change:

Key Equation: Conservation of angular momentum

$$\text{if } \sum \tau_{\text{ext}} = 0, \qquad \mathbf{L}_i = \mathbf{L}_f, \quad \text{i.e.} \quad I_i\omega_i = I_f\omega_f. \tag{11.8}$$

The last form is especially useful: if a body can rearrange its own mass, changing its moment of inertia, its angular speed must change to compensate. A figure skater pulling in her arms reduces I and spins faster; the same physics speeds up a collapsing star into a rapidly spinning pulsar, and shapes the flattened disks of galaxies and solar systems. Conservation of angular momentum is, quite literally, why the universe looks the way it does.

Two refinements make the law far more useful in practice, and both are easy to miss.

It holds component by component. Equation (11.8) is a vector statement, so each component of $\mathbf{L}$ is conserved separately. In particular, if the net external torque *about some particular axis* is zero, then the component of $\mathbf{L}$ along that axis is conserved, whatever the other components are doing. This is what rescues problems in which the system is plainly being pushed on from outside. When a lump of clay strikes a rod pivoted at one end, the pivot exerts a large external force during the collision, so linear momentum is emphatically *not* conserved, but that force acts *at* the pivot, with zero lever arm, so it exerts no torque about the pivot, and angular momentum about the pivot is conserved. Choosing the right axis converts an intractable problem into a one-line calculation.

Kinetic energy is generally not conserved along with it. Writing $K_{\text{rot}} = \frac{1}{2}I\omega^2$ in terms of the conserved quantity $L = I\omega$ gives

$$K_{\text{rot}} = \frac{L^2}{2I}, \tag{11.9}$$

so at fixed L, halving the moment of inertia *doubles* the kinetic energy. The skater who pulls in her arms ends up with more kinetic energy than she started with, and the energy is not free: to move her arms inward she must pull them against the outward tendency of masses traveling in a circle, and her muscles do exactly that work. Run the process the other way (letting the arms swing back out), and the energy is returned. In the opposite kind of process, where two rotating objects couple together and end up sharing a common ω, equation (11.9) with I *increasing* shows that kinetic energy is lost, exactly as it is in a perfectly inelastic collision. Angular momentum tells you the final angular speed; it never tells you that energy was conserved.

Example 11.4: Spin-up of a collapsing star

A star rotates once every 30 days. After a supernova, its core (of radius $10000\ \text{km}$) collapses to a neutron star of radius $3.0\ \text{km}$. Treating the core as a uniform sphere of unchanged mass, find the neutron star's rotation period.

Solution

Step 1: Conservation. No external torque acts, so angular momentum is conserved: $I_i\omega_i = I_f\omega_f$.

Step 2: Reduce it to radii and periods. For a uniform sphere $I = \frac{2}{5}MR^2$, and an angular speed of one turn per period is $\omega = 2\pi/T$. Substituting both,

$$\tfrac{2}{5}MR_i^2 \cdot \frac{2\pi}{T_i} = \tfrac{2}{5}MR_f^2 \cdot \frac{2\pi}{T_f}. \tag{11.4.1}$$

The mass is unchanged, so $\frac{2}{5}M$ and 2π appear on both sides and cancel, leaving $R_i^2/T_i = R_f^2/T_f$. Solve for T_f: multiply both sides by T_fT_i and divide both sides by R_i^2:

$$T_f = T_i \frac{R_f^2}{R_i^2} = T_i \left(\frac{R_f}{R_i} \right)^2 . \tag{11.4.2}$$

Shrinking the radius by some factor shrinks the period by the *square* of that factor.

Step 3: Substitute. Convert the period to seconds, $T_i = 30 \text{ days} \times 86\ 400 \text{ s/day} = 2.59 \times 10^6 \text{ s}$. Both radii are in kilometers, so their ratio needs no conversion: $3.0/10\ 000 = 3.0 \times 10^{-4}$. Squaring it squares both the 3.0 and the power of ten, $(3.0)^2 = 9.0$ and $(10^{-4})^2 = 10^{-8}$:

$$T_f = (2.59 \times 10^6 \text{ s})(3.0 \times 10^{-4})^2 = (2.59 \times 10^6)(9.0 \times 10^{-8}) \text{ s} \approx \boxed{0.23 \text{ s}}. \tag{11.4.3}$$

(Multiplying the powers of ten, $10^6 \times 10^{-8} = 10^{-2}$, and $2.59 \times 9.0 = 23.3$, giving $23.3 \times 10^{-2} = 0.233$.)

A star that took a month to turn now spins more than four times a second ($1/0.23 \approx 4.3$): a pulsar. By equation (11.9) its rotational kinetic energy has risen by the same factor of about 10^7 that its moment of inertia fell; that energy came from gravity doing work as the core collapsed.

Example 11.5: A student on a turntable

A horizontal disk-shaped platform ($M = 100 \text{ kg}$, $R = 2.0 \text{ m}$) rotates freely on a frictionless axle. A 60 kg student stands at the rim while the system turns at 2.0 rad/s. The student walks inward to a point 0.50 m from the center. Find the new angular speed.

Solution

Step 1: Conservation. With no external torque about the vertical axle, the total angular momentum of platform plus student is conserved: $I_i \omega_i = I_f \omega_f$. Only the student's position changes, so only the student's contribution to I changes.

Step 2: The moments of inertia. The disk contributes $I_{\text{disk}} = \frac{1}{2} MR^2 = \frac{1}{2}(100 \text{ kg})(2.0 \text{ m})^2 = 200 \text{ kg} \cdot \text{m}^2$ throughout. The student counts as a point mass, mr^2, where r is the student's distance from the axle. Moments of inertia about the same axis simply add:

$$\begin{aligned} I_i &= 200 + (60)(2.0)^2 = 200 + 240 = 440 \text{ kg} \cdot \text{m}^2, \\ I_f &= 200 + (60)(0.50)^2 = 200 + 15 = 215 \text{ kg} \cdot \text{m}^2. \end{aligned} \tag{11.4.4}$$

Moving from the rim to 0.50 m reduces the student's distance by a factor of 4 and so reduces the student's mr^2 by a factor of 16, from 240 to $15 \text{ kg} \cdot \text{m}^2$.

Step 3: Solve for ω_f. Divide both sides of $I_i \omega_i = I_f \omega_f$ by I_f:

$$\omega_f = \frac{I_i \omega_i}{I_f} = \frac{(440 \text{ kg} \cdot \text{m}^2)(2.0 \text{ rad/s})}{215 \text{ kg} \cdot \text{m}^2} \approx \boxed{4.1 \text{ rad/s}}. \tag{11.4.5}$$

Moving inward lowers the moment of inertia, so the platform spins up: the skater effect on a larger scale.

Energy is not conserved. The kinetic energy rises from $\frac{1}{2}(440)(2.0)^2 = 880 \text{ J}$ to $\frac{1}{2}(215)(4.09)^2 = 1.80 \times 10^3 \text{ J}$; the extra 920 J is work done by the student in walking inward against the outward push the student feels. Angular momentum, by contrast, is the same before and after: $(440)(2.0) = 880$ and $(215)(4.09) = 880 \text{ kg} \cdot \text{m}^2/\text{s}$. ✓

Checkpoint 11.4

You stand at rest on a frictionless turntable, holding a ball.

a. You throw the ball horizontally, aiming so that its line of flight passes directly over the turntable's axis. Do you begin to spin? Explain using equation (11.6).
b. You repeat the experiment, throwing the ball just as hard but off to one side. What happens now, and which way do you turn?

c. In part (b), is the linear momentum of the you-plus-ball system conserved? Is its angular momentum about the axis conserved? Explain why the two answers can differ.

Answer

With angular momentum in hand, the correspondence between translation and rotation that has run through both rotation chapters is complete. Table 11.2 adds the last few rows to Table 11.1.

Table 11.2. Completing the translational–rotational correspondence begun in Table 11.1.

Translation	Rotation
momentum $\mathbf{p} = m\mathbf{v}$	angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, or $L = I\omega$
$\sum \mathbf{F} = d\mathbf{p}/dt$	$\sum \tau = d\mathbf{L}/dt$
$\mathbf{p}$ conserved if $\sum \mathbf{F}_{\text{ext}} = 0$	$\mathbf{L}$ conserved if $\sum \tau_{\text{ext}} = 0$
$K = p^2/2m$	$K_{\text{rot}} = L^2/2I$

Answers to Checkpoints

Checkpoint 11.4. (a) No. The ball's line of flight passes through the axis, so $r_\perp = 0$ and the ball carries zero angular momentum about the axis. The total started at zero and stays zero, so you do not spin. (The axle stops you from recoiling across the room.) (b) Now the ball's line of flight misses the axis, so it carries angular momentum $r_\perp mv$ about the axis. To keep the total at zero, you and the turntable must rotate the opposite way: if the ball circulates counterclockwise about the axis as seen from above, you turn clockwise. (c) Linear momentum is *not* conserved: the axle pushes horizontally on the turntable, an external force that stops you from recoiling. Angular momentum about the axis *is* conserved, because that axle force acts at the axis, where its lever arm is zero and so its torque about the axis vanishes.

This page titled 11.4: Conservation of Angular Momentum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.5: Precession

A spinning top does something that seems to defy gravity: instead of toppling over, its axis sweeps slowly around a vertical circle. This motion is **precession**, and it follows directly from $\sum \tau = d\mathbf{L}/dt$ once we take the vector nature of angular momentum seriously.

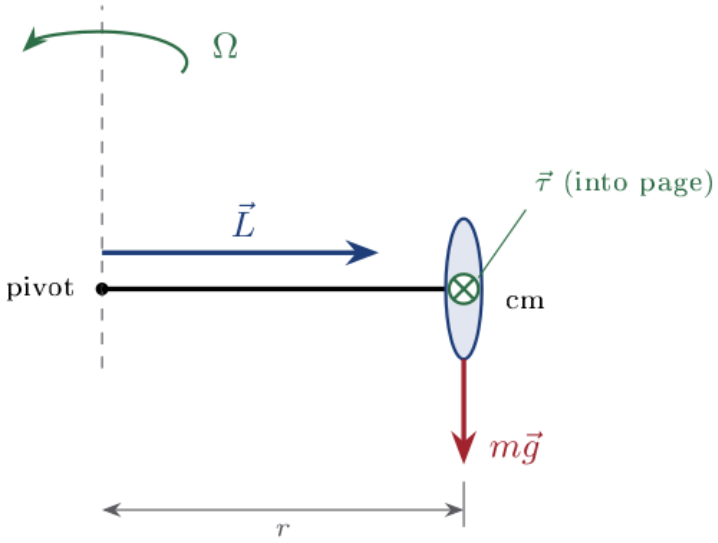


Figure 11.1. A gyroscope precesses. Gravity exerts a torque $\tau = \mathbf{r} \times m\mathbf{g}$ perpendicular to the angular momentum $\mathbf{L}$ (here into the page), which swings $\mathbf{L}$ sideways rather than tipping it down. The axle sweeps around the vertical at rate $\Omega = mgr/(I\omega)$.

•Figure description

A gyroscope on a pivot. The spin angular momentum L points along the axle, gravity acts at the center of mass a distance r from the pivot, and the resulting torque is perpendicular to L, causing the axle to precess at rate capital Omega.

Consider a gyroscope: a spinning disk on a horizontal axle pivoted at one end (Figure 11.1). Its angular momentum $\mathbf{L}$ points along the axle, horizontally. Gravity, acting at the center of mass a distance r from the pivot, produces a torque $\tau = \mathbf{r} \times m\mathbf{g}$ of magnitude mgr, directed *horizontally and perpendicular to* $\mathbf{L}$. By equation (11.7), this torque changes $\mathbf{L}$ in its own direction, not by tipping $\mathbf{L}$ down, but by swinging it sideways.

The rate follows in three lines. In a short time dt the angular momentum changes by $d\mathbf{L} = \tau\ dt$, a small vector of length $\tau\ dt$ lying perpendicular to $\mathbf{L}$. Adding a perpendicular increment to a vector rotates it without changing its length, through an angle

$$d\phi = \frac{|d\mathbf{L}|}{L} = \frac{\tau\, dt}{L}. \tag{11.5.1}$$

The tip of $\mathbf{L}$ therefore travels in a horizontal circle at the angular rate $\Omega = d\phi/dt$:

Key Equation: Precession angular velocity

$$\Omega = \frac{\tau}{L} = \frac{mgr}{I\omega}. \tag{11.10}$$

Notice that a faster spin (larger $L = I\omega$) gives a *slower* precession, which is why a rapidly spinning top precesses lazily, and why a gyroscope's stiff, steady axis makes it the heart of navigational instruments. Two remarks on the result. First, it assumes $\Omega \ll \omega$, so that essentially all of the body's angular momentum lies along the spin axis; a slow top also nods up and down as it precesses, a wobble called nutation that this treatment leaves out. Second, although the figure shows a horizontal axle, equation (11.10) holds at any tilt: tipping the axle by an angle ϕ from the vertical reduces the torque by a factor $\sin\phi$, but it also reduces the horizontal part of $\mathbf{L}$ that has to be swung around by the same factor, and the two cancel.

Example 11.6: A bicycle wheel as a gyroscope

A bicycle wheel of mass $2.0\ \mathrm{kg}$ and moment of inertia $0.14\ \mathrm{kg\cdot m^2}$ is mounted on a light axle and spun up to $30\ \mathrm{rad/s}$. One end of the axle is then set on a pivot, leaving the wheel's center $0.30\ \mathrm{m}$ from the pivot with the axle horizontal. Find the precession angular velocity and the time for one full circuit of the axle.

Solution

Directly from equation (11.10). The numerator mgr is the gravitational torque about the pivot (the weight mg acting at lever arm r); the denominator $I\omega$ is the spin angular momentum:

$$\Omega = \frac{mgr}{I\omega} = \frac{(2.0\,\mathrm{kg})(9.8\,\mathrm{m/s^2})(0.30\,\mathrm{m})}{(0.14\,\mathrm{kg\cdot m^2})(30\,\mathrm{rad/s})} = \frac{5.88\,\mathrm{N\cdot m}}{4.2\,\mathrm{kg\cdot m^2/s}} = \boxed{1.4\,\mathrm{rad/s}}. \tag{11.5.2}$$

The units reduce correctly: $\mathrm{N\cdot m} = \mathrm{kg\cdot m^2/s^2}$, and dividing by $\mathrm{kg\cdot m^2/s}$ leaves $\mathrm{s^{-1}}$, an angular speed.

One full circuit is 2π radians, so at angular speed Ω it takes

$$T_{\mathrm{prec}} = \frac{2\pi}{\Omega} = \frac{2\pi}{1.4\,\mathrm{rad/s}} \approx \boxed{4.5\,\mathrm{s}}. \tag{11.5.3}$$

The approximation behind equation (11.10) is comfortably satisfied: $\Omega/\omega = 1.4/30 \approx 0.05$, so the wheel turns about twenty times on its axle for each slow sweep around the pivot. Because ω sits in the denominator, a faster spin gives a *slower* precession: double the spin and the axle takes twice as long to go around. Note also what does *not* appear in the answer: nothing about how the wheel is released. Let go of a non-spinning wheel in this position and it simply falls.

Checkpoint 11.5

A toy gyroscope is precessing counterclockwise as seen from above.

a. You stop it, reverse the direction of its spin, and release it again from the same position. Which way does it precess now?
b. You then move the wheel outward along the axle, doubling r, without changing anything else. What happens to Ω?
c. Equation (11.10) predicts $\Omega \to \infty$ as $\omega \to 0$. What actually happens to a real top as friction brings its spin down, and why does the equation stop describing it?

Answer

Answers to Checkpoints

Checkpoint 11.5. (a) Clockwise. Reversing the spin reverses $\mathbf{L}$, while the gravitational torque, which depends only on where the wheel hangs, is unchanged. Since $d\mathbf{L} = \tau\ dt$ still points the same way, the reversed $\mathbf{L}$ swings around in the opposite sense. (b) The torque mgr doubles, and nothing else in equation (11.10) changes, so Ω doubles. (c) A real top does not precess infinitely fast. As its spin slows, it precesses faster, then wobbles (nutates) more and more wildly, and finally falls over. The equation assumes the spin angular momentum is so large that the slow precession adds negligibly to it ($\Omega \ll \omega$); when ω becomes small, that assumption fails, and the equation no longer describes the motion.

This page titled 11.5: Precession is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.S: Rotation II (Summary)

Summary

- **Rotational kinetic energy** is $K_{\text{rot}} = \frac{1}{2} I\omega^2$. A body rolling without slipping carries both kinds, $K = \frac{1}{2} m v_{\text{cm}}^2 + \frac{1}{2} I_{\text{cm}} \omega^2$.
- **Static friction** does no work on a rolling body, because the contact point does not move, so energy conservation applies: a body with $I_{\text{cm}} = \beta m R^2$ that rolls down a height h reaches $v_{\text{cm}} = \sqrt{2gh/(1+\beta)}$.
- A single torque does **work** $W = \tau\ \Delta\theta$ and delivers **power** $P = \tau\omega$; the *net* work of all torques obeys $W_{\text{net}} = \Delta K_{\text{rot}}$. Rotation parallels translation quantity for quantity (Tables 11.1 and 11.2).
- **Angular momentum** is $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ for a particle (equivalently $L = r_\perp mv$) and $L = I\omega$ for a rigid body. It is always defined about a chosen point or axis. The net torque equals its rate of change, $\sum \tau = d\mathbf{L}/dt$.
- **Conservation of angular momentum**: with no external torque, $I_i\omega_i = I_f\omega_f$. It holds component by component, so L about an axis is conserved whenever the net torque about that axis vanishes, even when linear momentum is not conserved. Reducing I raises ω (the skater effect, the pulsar, the galactic disk), and since $K_{\text{rot}} = L^2/2I$, it raises the kinetic energy too.
- A spinning gyroscope **precesses** at $\Omega = mgr/(I\omega)$: a gravitational torque perpendicular to $\mathbf{L}$ swings the axis around instead of toppling it.

Key Equations

Rotational kinetic energy

$$K_{\text{rot}} = \tfrac{1}{2} I\omega^2 \tag{11.S.1}$$

Kinetic energy of a rolling body

$$K = \tfrac{1}{2} m v_{\text{cm}}^2 + \tfrac{1}{2} I_{\text{cm}} \omega^2 \tag{11.S.2}$$

Work done by a torque

$$W = \tau\,\Delta\theta, \qquad W_{\text{net}} = \Delta K_{\text{rot}} = \tfrac{1}{2} I\omega^2 - \tfrac{1}{2} I\omega_0^2 \tag{11.S.3}$$

Power delivered by a torque

$$P = \tau\omega \tag{11.S.4}$$

Angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad |\mathbf{L}| = rmv\sin\theta; \qquad L = I\omega \text{ (rigid body)} \tag{11.S.5}$$

Newton's second law (angular momentum form)

$$\sum \tau = \frac{d\mathbf{L}}{dt} \tag{11.S.6}$$

Conservation of angular momentum

$$\text{if } \sum \tau_{\text{ext}} = 0, \qquad \mathbf{L}_i = \mathbf{L}_f, \quad \text{i.e.} \quad I_i\omega_i = I_f\omega_f \tag{11.S.7}$$

Precession angular velocity

$$\Omega = \frac{\tau}{L} = \frac{mgr}{I\omega} \tag{11.S.8}$$

This page titled 11.S: Rotation II (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

11.E: Rotation II (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 11.1 — Rotational Kinetic Energy

11.1 ★ A flywheel has a moment of inertia of $2.5\ \text{kg}\cdot\text{m}^2$ and spins at $40\ \text{rad/s}$. Find its rotational kinetic energy.

11.2 ★★ *[Real-world: energy storage.]* A flywheel (solid disk, $120\ \text{kg}$, radius $0.50\ \text{m}$) stores energy by spinning at $3000\ \text{rpm}$.

a. Find its moment of inertia.
b. How much energy does it store?

11.3 ★★ *[Conceptual.]* Two solid spheres have the same mass, but one has twice the radius of the other. Spun at the same angular velocity, which stores more rotational kinetic energy, and by what factor?

11.4 ★★ A solid disk of mass $3.0\ \text{kg}$ and radius $0.20\ \text{m}$ rolls without slipping at $4.0\ \text{m/s}$.

a. Find its translational and rotational kinetic energies.
b. What fraction of its total kinetic energy is rotational?

11.5 ★★ A solid sphere ($I = \frac{2}{5}mR^2$) is released from rest at the top of a ramp of height $1.2\ \text{m}$ and rolls down without slipping. Use energy conservation to find its center-of-mass speed at the bottom. *(Use $v_{\text{cm}} = R\omega$.)*

11.6 ★★ *[Conceptual.]* A hoop, a solid disk, and a solid sphere of equal mass and radius are released together from the top of a ramp and roll down without slipping.

a. In what order do they reach the bottom? Explain in terms of how each one's moment of inertia splits the energy between translation and rotation.
b. Check your ordering against equation (10.15), and explain why the masses and radii drop out of that result.
c. Which of the three needs the roughest ramp to keep from slipping?

Section 11.2 — Rotational Work

11.7 ★ A constant torque of $15\ \text{N}\cdot\text{m}$ turns a wheel through $20\ \text{rad}$. How much work does it do?

11.8 ★★ A motor applies a constant $25\ \text{N}\cdot\text{m}$ torque to a wheel ($I = 0.80\ \text{kg}\cdot\text{m}^2$), initially at rest, through $12\ \text{rad}$.

a. How much work does the motor do?
b. Using the work–kinetic energy theorem, find the wheel's final angular speed.

11.9 ★★★ *[Real-world: sports.]* A solid disk (a discus idealization) of mass $2.0\ \text{kg}$ and radius $0.11\ \text{m}$ is spun up from rest to $45\ \text{rad/s}$.

a. How much rotational kinetic energy does it have?
b. If this is achieved over an angular displacement of $6.0\ \text{rad}$, what average torque was applied?

11.10 ★ *[Conceptual.]* The same constant torque acts on two wheels of equal moment of inertia, but wheel A is allowed to turn through twice the angle that wheel B does.

a. On which wheel is more work done? Explain.
b. Which ends up with the greater rotational kinetic energy, and which with the greater angular speed? Justify using the rotational work–kinetic energy theorem.

Section 11.3 — Angular Momentum

11.11 ★ A $0.50\ \text{kg}$ ball moves at $4.0\ \text{m/s}$ along a line that passes $2.0\ \text{m}$ from a fixed point O. Find the magnitude of its angular momentum about O.

11.12 ★ A solid cylinder ($I = \frac{1}{2}MR^2$) of mass $2.0\ \text{kg}$ and radius $0.15\ \text{m}$ spins about its central axis at $20\ \text{rad/s}$. Find its angular momentum.

11.13 ★★ A solid disk of moment of inertia $0.40\ \text{kg}\cdot\text{m}^2$ spins at $15\ \text{rad/s}$.

a. Find its angular momentum.
b. What constant torque would stop it in $3.0\ \text{s}$? *(Use $\tau = dL/dt$.)*

11.14 ★★ *[Conceptual.]* Explain why a spinning bicycle wheel held by its axle is hard to tilt, while the same wheel not spinning tilts easily. Frame your answer in terms of angular momentum and torque.

Section 11.4 — Conservation of Angular Momentum

11.15 ★ *[Conceptual.]* A figure skater spins with her arms outstretched, then pulls them in close to her body.

a. What happens to her rate of spin, and why?
b. Her rotational kinetic energy increases. Where does that extra energy come from?

11.16 ★★ A merry-go-round (a uniform disk, $I = 500\ \text{kg}\cdot\text{m}^2$) turns freely at $1.5\ \text{rad/s}$. A $40\ \text{kg}$ child, initially at the center, walks out to the rim at $2.0\ \text{m}$. Find the new angular speed. *(Treat the child as a point mass.)*

11.17 ★★ *[Real-world: astronomy.]* A spinning cloud of gas contracts to one-tenth of its original radius with no external torque. Treating it as a uniform sphere of unchanged mass, by what factor does its angular speed increase? By what factor does its rotational kinetic energy increase?

11.18 ★★★ A $0.20\ \text{kg}$ puck moves in a circle of radius $0.80\ \text{m}$ on a frictionless table at $3.0\ \text{m/s}$, held by a string through a hole in the center. Find its angular momentum. If the string is pulled to shrink the radius to $0.40\ \text{m}$, find the new speed using conservation of angular momentum. Why does the string tension exert no torque about the hole?

11.19 ★★★ A $2.0\ \text{kg}$ disk of radius $0.25\ \text{m}$ spins at $12\ \text{rad/s}$. A second, non-spinning disk of the same size and mass is dropped onto it and the two spin together (a rotational "perfectly inelastic collision").

a. Find the common final angular speed.
b. How much rotational kinetic energy is lost?

Section 11.5 — Precession

11.20 ★ *[Conceptual.]* Using $\Omega = mgr/(I\omega)$, explain why a rapidly spinning top precesses slowly, and predict what happens to the precession as friction gradually slows the spin.

11.21 ★★ A toy gyroscope's wheel has moment of inertia $2.0 \times 10^{-4}\ \text{kg}\cdot\text{m}^2$ and spins at $150\ \text{rad/s}$. Its center of mass is $0.040\ \text{m}$ from the pivot, and its mass is $0.060\ \text{kg}$. Find the precession angular velocity.

11.22 ★★ A gyroscope precesses once every $6.0\ \text{s}$ while its wheel ($I = 5.0 \times 10^{-3}\ \text{kg}\cdot\text{m}^2$, $m = 0.25\ \text{kg}$) spins with its center of mass $0.080\ \text{m}$ from the pivot. Find the spin angular velocity of the wheel, and verify that $\Omega \ll \omega$ as the derivation of equation (11.10) assumes.

11.23 ★★★ *[Real-world: engineering.]* A flywheel-based attitude sensor uses a disk of mass $0.40\ \text{kg}$ and radius $0.050\ \text{m}$ spinning about a horizontal axle pivoted $0.020\ \text{m}$ from the disk's center. The design requires the precession period to be no shorter than $30\ \text{s}$.

a. What is the minimum spin rate, in rad/s and in rev/min?

b. Bearing friction slowly reduces the spin rate. Does the precession period grow or shrink, and by what factor after the spin rate has dropped by 10%?

Challenge Problems

11.24 ★★★ A uniform rod of mass $2.0\ \text{kg}$ and length $1.0\ \text{m}$ is pivoted at one end on a frictionless horizontal axle and released from rest in the horizontal position. (Take $I_{\text{end}} = \frac{1}{3}ML^2$.)

a. Find the rod's angular acceleration at the instant of release.
b. Find the linear acceleration of the free end at that instant, and compare it with g.
c. Using energy conservation, find the rod's angular speed when it reaches the vertical position.
d. How fast is the free end moving at that point?

11.25 ★★★ *[Real-world: engineering.]* A flywheel (solid disk, $M = 50\ \text{kg}$, $R = 0.40\ \text{m}$) spins at $1200\ \text{rpm}$. A constant friction torque brings it to rest in $90\ \text{s}$.

a. Find the flywheel's moment of inertia and its initial rotational kinetic energy.
b. Find the (constant) friction torque.
c. Through how many revolutions does it turn before stopping?
d. What is the average power dissipated by friction?

11.26 ★★★ A $50\ \text{g}$ lump of putty moving at $12\ \text{m/s}$ strikes the free end of a uniform rod ($M = 0.50\ \text{kg}$, $L = 0.80\ \text{m}$) pivoted at its other end, hitting perpendicular to the rod and sticking. (Use $I_{\text{rod}} = \frac{1}{3}ML^2$.)

a. Find the angular momentum of the putty about the pivot just before impact.
b. Find the moment of inertia of the rod-plus-putty system.
c. Find the angular speed immediately after the collision.
d. Through what maximum angle does the rod swing up afterward?
e. Linear momentum is clearly not conserved in this collision. Explain why angular momentum about the pivot nevertheless is.

11.27 ★★★ A student sits at rest on a frictionless rotating stool (combined moment of inertia $3.0\ \text{kg}\cdot\text{m}^2$ about the vertical axis), holding a spinning bicycle wheel ($I = 0.40\ \text{kg}\cdot\text{m}^2$) whose axis points straight up; the wheel spins at $15\ \text{rad/s}$.

a. What is the total angular momentum about the vertical axis?
b. The student flips the wheel so its axis points straight down. How fast, and in which sense, does the stool now rotate?
c. Starting again from the original state, the student instead turns the wheel's axis to horizontal. How fast does the stool rotate then?
d. Why is the vertical component of angular momentum conserved here?

This page titled 11.E: Rotation II (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12: STATIC EQUILIBRIUM AND ELASTICITY

Not everything in mechanics is about motion. Bridges, buildings, bones, and bookshelves all have one job: to stay put under load. A structure that holds still is in **static equilibrium**, and analyzing it is one of the most practical applications of everything we have built so far. The tools are already in hand (forces from Chapter 5 and torques from Chapter 10), but now we demand that they all cancel, so that nothing accelerates and nothing rotates.

Real materials, of course, are not perfectly rigid: push or pull hard enough and they stretch, compress, bend, or break. The second half of this chapter takes up **elasticity**: how solids deform under stress and how much force they can take. Together, equilibrium and elasticity are the foundation of statics and of structural engineering.

Learning Objectives

- State the two conditions for static equilibrium and apply them to extended objects.
- Solve for unknown forces and torques on beams, ladders, and supports.
- Locate and use the center of gravity of an object.
- Predict when an object will tip, using the center of gravity and the base of support.
- Relate stress and strain through the elastic moduli (Young's, shear, and bulk).
- Compute deformations and connect an object's stiffness to the material it is made of.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

12.1: Equilibrium
12.2: Center of Gravity and Stability
12.3: Elasticity
12.S: Static Equilibrium and Elasticity (Summary)
12.E: Static Equilibrium and Elasticity (Problems)

This page titled 12: Static Equilibrium and Elasticity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12.1: Equilibrium

An object is in **static equilibrium** when it is at rest and stays at rest. Two conditions must hold at once. First, the linear momentum of its center of mass is constant, meaning the net force is zero. Second, its angular momentum about any point is constant, meaning the net torque is zero:

Key Equation: Conditions for static equilibrium

$$\sum \mathbf{F} = 0 \qquad \text{and} \qquad \sum \tau = 0. \tag{12.1}$$

The force condition is really one equation per direction ($\sum F_x = 0$, $\sum F_y = 0$). The torque condition comes with a feature that makes statics problems tractable, and it is worth seeing exactly where it comes from. Compare the net torque about two different points O and P. Every position vector shifts by the same fixed displacement $\mathbf{r}_{O\to P}$, so

$$\sum \tau_P = \sum \tau_O + \mathbf{r}_{O\to P} \times \sum \mathbf{F}. \tag{12.1.1}$$

If the force condition already holds, the last term vanishes and *the net torque is the same about every point*. So once we have imposed $\sum \mathbf{F} = 0$, checking $\sum \tau = 0$ about one point guarantees it about all of them, and we are free to pick whichever point makes the algebra easiest. Choosing the point where an unknown force acts gives that force zero lever arm, dropping it out of the torque equation entirely. That choice is often the whole solution.

A reliable strategy: draw a free-body diagram, write $\sum F_x = 0$ and $\sum F_y = 0$, then write $\sum \tau = 0$ about a well-chosen point. Two habits matter here. First, the free-body diagram must be an *extended* one: unlike the point-particle diagrams of Chapter 5, where every arrow started at the same dot, each force must now be drawn at the location where it actually acts, because that location is what sets its lever arm. Second, fix a sign convention for torques (we take counterclockwise as positive) and stay with it. In simple problems it is quicker to set the clockwise torques equal to the counterclockwise ones, which is what the examples below do, but the signed form is what scales to complicated cases. The weight of an object is taken to act at its center of gravity (§12.2).

Note: How many unknowns can statics handle?

In two dimensions the equilibrium conditions supply exactly three independent equations: $\sum F_x = 0$, $\sum F_y = 0$, and $\sum \tau = 0$. Writing the torque equation about a second point does not give a fourth: as we just showed, it is the same equation in disguise. So statics alone can determine at most *three* unknowns in a planar problem.

Structures with more unknowns than that are called **statically indeterminate**. A beam resting on three supports, or a table on four legs, is a familiar example: the equations cannot tell you how the load divides among the supports, because the answer depends on how much each support deflects. Solving such problems requires the elasticity of §12.3 in addition to the equilibrium conditions, and is a central topic of a mechanics-of-materials course. Every problem in this chapter is deliberately determinate.

Checkpoint 12.1

Two blocks, of masses m and M, hang from cords wrapped around two different-sized disks fixed on a common axle (radii r and R respectively). Someone claims the system balances when $mr^2 = MR^2$, reasoning from the moment of inertia $I = mr^2$. Why is that wrong, and what is the actual condition for static equilibrium? *(Hint: equilibrium is about torque, not rotational inertia.)*

Answer

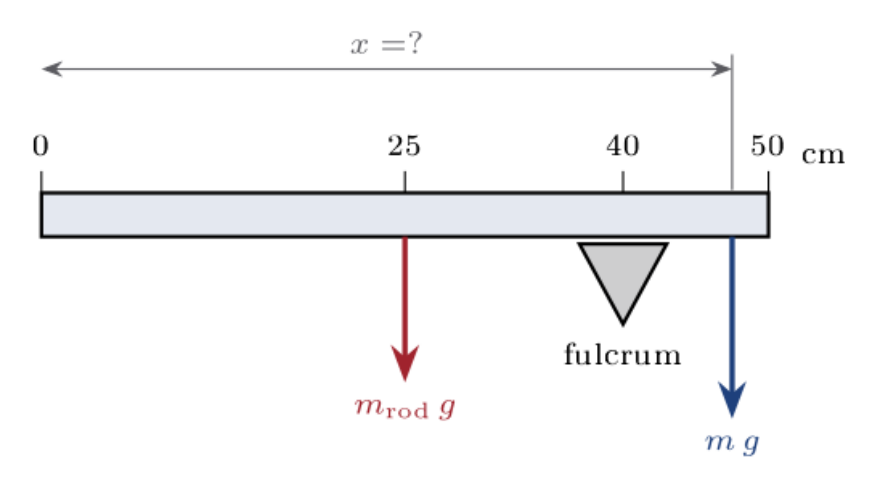
Checkpoint 12.2

A horizontal beam is pinned to a wall at its left end and held up by a cable at its right end; a load hangs somewhere in between. You want the cable tension, and you want it without ever solving for the pin force. About which point should you take torques, and why does that choice eliminate the pin force rather than merely simplify it?

Answer

Example 12.1: Balancing a rod on a fulcrum

A uniform rod of length $50\ \text{cm}$ and mass $0.20\ \text{kg}$ rests on a fulcrum $40\ \text{cm}$ from its left end. How far from the left end should a $0.60\ \text{kg}$ mass be hung to balance the rod?



•Figure description
A rod balanced on a fulcrum, with its weight acting at its center of mass and a hanging mass placed an unknown distance x from the left end.

Solution

Step 1: Locate the forces. The rod's weight acts at its center, $25\ \text{cm}$ from the left, that is, $40 - 25 = 15\ \text{cm}$ to the *left* of the fulcrum. That weight tries to turn the rod counterclockwise, so the hung mass must sit to the *right* of the fulcrum to balance it. Let x be the mass's distance from the left end; its lever arm about the fulcrum is then $x - 0.40$ (in meters).

Step 2: Balance torques about the fulcrum. Choosing the fulcrum as the pivot point means the unknown support force there has zero lever arm and drops out. Counterclockwise torque equals clockwise torque:

$$m_{\text{rod}}\, g\,(0.40 - 0.25) = m\, g\,(x - 0.40). \tag{12.1.2}$$

Step 3: Solve. The g appears on both sides, so divide it out. Then divide both sides by 0.60 and add 0.40:

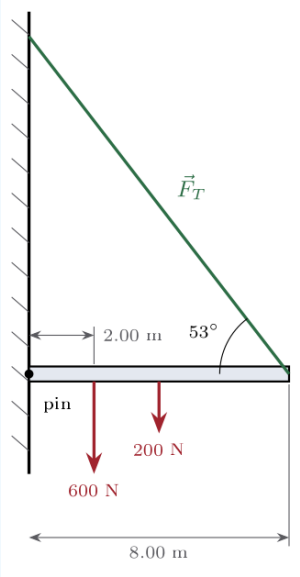
$$(0.20)(0.15) = (0.60)(x - 0.40) \implies x - 0.40 = \frac{0.030}{0.60} = 0.05 \implies x = \boxed{0.45\,\text{m}}, \tag{12.1.3}$$

or $45\ \text{cm}$ from the left end.

Check. The hung mass is three times the rod's mass, so it needs one-third the lever arm: $15\ \text{cm}/3 = 5\ \text{cm}$ to the right of the fulcrum, which is $45\ \text{cm}$ from the left end. It also lies on the rod ($45\ \text{cm} < 50\ \text{cm}$), so the arrangement is possible.

Example 12.2: Beam supported by a cable

A uniform horizontal beam $8.00\ \text{m}$ long and weighing $200\ \text{N}$ is attached to a wall by a pin. Its far end is held by a cable making a 53° angle with the beam. A person weighing $600\ \text{N}$ stands $2.00\ \text{m}$ from the wall. Find the cable tension and the force the wall exerts on the beam.



•Figure description
A horizontal beam attached to a wall by a pin, supported by a cable at 53 degrees. The beam's own 200 newton weight acts at its center and a 600 newton load stands 2 meters from the wall.

Solution

Step 1: Torques about the pin. Take torques about the pin, which eliminates the unknown wall force (its lever arm is zero). Resolve the cable tension into components along and perpendicular to the beam. The component along the beam, $F_T \cos 53^\circ$, points straight at the pin and has no lever arm; only the perpendicular component, $F_T \sin 53^\circ$, produces a torque, acting at the far end ($8.00\ \text{m}$). The two weights act downward, the beam's at its center ($4.00\ \text{m}$) and the person's at $2.00\ \text{m}$. The cable's torque (counterclockwise) balances the two weights' torques (clockwise):

$$F_T \sin 53^\circ\,(8.00\ \text{m}) = (200\ \text{N})(4.00\ \text{m}) + (600\ \text{N})(2.00\ \text{m}) = 2000\ \text{N}\cdot\text{m}. \tag{12.1.4}$$

Divide both sides by $(8.00\ \text{m})\sin 53^\circ$:

$$F_T = \frac{2000\ \text{N}\cdot\text{m}}{(8.00\ \text{m})(0.799)} \approx \boxed{313\,\text{N}}. \tag{12.1.5}$$

Step 2: The wall force. Now the force conditions give the wall force. A pin can push in any direction (it is *not* restricted to pushing along the beam, a common first guess), so we treat its force as two unknown components, horizontal $F_{\text{wall},x}$ and vertical $F_{\text{wall},y}$, and recover the direction at the end.

Horizontally, the cable's horizontal component $F_T \cos 53^\circ$ pulls the beam toward the wall, so the pin must push outward by the same amount:

$$F_{\text{wall},x} = F_T \cos 53^\circ = (313\,\text{N})(0.602) \approx 188\,\text{N}. \tag{12.1.6}$$

Vertically, upward forces balance downward ones: $F_{\text{wall},y} + F_T \sin 53^\circ = 200 + 600$. From Step 1, $F_T \sin 53^\circ = 2000/8.00 = 250\ \text{N}$ exactly, so subtracting it from both sides,

$$F_{\text{wall},y} = 800 - 250 = 550\,\text{N}. \tag{12.1.7}$$

Step 3: Magnitude and direction. The two components are perpendicular, so they combine by the Pythagorean theorem, and

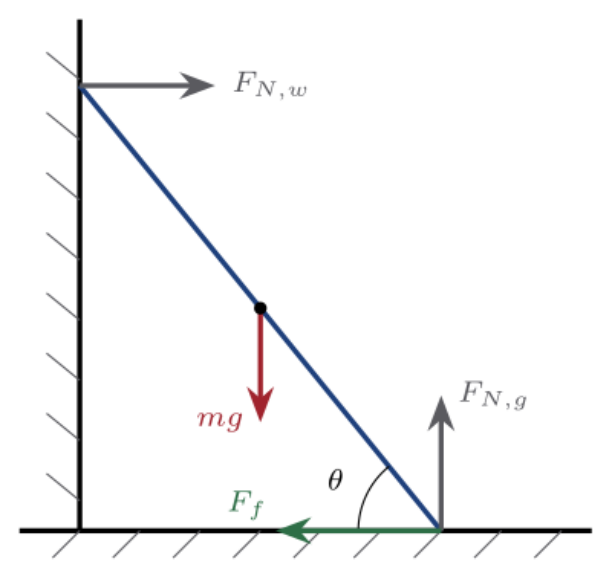
the angle above the horizontal has $F_{\text{wall},y}$ opposite and $F_{\text{wall},x}$ adjacent:

$$F_{\text{wall}} = \sqrt{F_{\text{wall},x}^2 + F_{\text{wall},y}^2} = \sqrt{(188)^2 + (550)^2}\,\text{N} \approx \boxed{581\,\text{N}},$$
$$\theta = \arctan\frac{F_{\text{wall},y}}{F_{\text{wall},x}} = \arctan\frac{550}{188} \approx \boxed{71^\circ \text{ above horizontal}}. \tag{12.1.8}$$

Check. Any point can serve as the pivot in equilibrium. Take torques about the cable's end of the beam instead: now F_T drops out, and $F_{\text{wall},y}$ (at 8.00 m) must balance the beam's weight (at 4.00 m from that end) and the person's (at 6.00 m): $8.00\ V = (200)(4.00) + (600)(6.00) = 4400$, so $F_{\text{wall},y} = 550$ N in one step. ✓

Example 12.3: A ladder against a smooth wall

A uniform ladder of length ℓ and mass m leans against a frictionless vertical wall. The coefficient of static friction with the ground is $\mu_s = 0.40$. Find the minimum angle with the ground at which the ladder will not slip.



•Figure description
A ladder leaning against a smooth wall at angle theta to the ground, showing the weight at its center, the normal force from the wall, and the normal force and friction at the ground.

Solution

Step 1: The forces. Four forces act: the ladder's weight mg at its center; the ground's normal $F_{N,g}$ (up) and friction F_f (toward the wall, since the base tends to slide out); and the wall's normal $F_{N,w}$ (away from the wall; no friction, since the wall is smooth).

Step 2: Force conditions. Vertically, only $F_{N,g}$ and the weight act; horizontally, only F_f and $F_{N,w}$:

$$F_{N,g} = mg, \qquad F_f = F_{N,w}. \tag{12.1.9}$$

Step 3: Torques about the base. Choosing the base as the pivot eliminates both $F_{N,g}$ and F_f, which act there. For each remaining force the lever arm is the perpendicular distance from the base to the force's line of action. The wall's normal is horizontal and acts at the top, so its lever arm is the *height* of the top, $\ell\sin\theta$. The weight is vertical and acts at the center, so its lever arm is the *horizontal* distance to the center, $\frac{\ell}{2}\cos\theta$. Balancing,

$$F_{N,w}\,(\ell\sin\theta) = mg\,\tfrac{\ell}{2}\cos\theta. \tag{12.1.10}$$

Divide both sides by $\ell\sin\theta$; the length cancels, and $\cos\theta/\sin\theta = 1/\tan\theta$:

$$F_{N,w} = \frac{mg}{2}\cdot\frac{\cos\theta}{\sin\theta} = \frac{mg}{2\tan\theta}. \tag{12.1.11}$$

Step 4: The slipping condition. The ladder is on the verge of slipping when friction reaches its maximum, $F_f = \mu_s F_{N,g} = \mu_s mg$. Setting $F_f = F_{N,w}$ and canceling mg from both sides:

$$\mu_s mg = \frac{mg}{2\tan\theta} \implies \mu_s = \frac{1}{2\tan\theta}. \tag{12.1.12}$$

Multiply both sides by $\tan\theta$ and divide by μ_s:

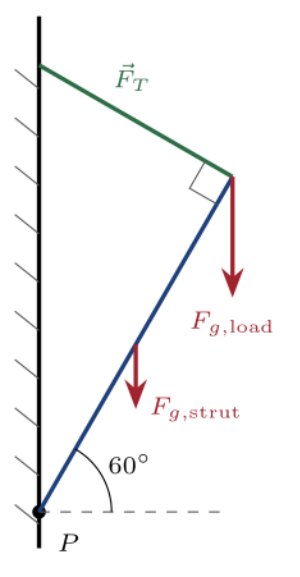
$$\tan\theta = \frac{1}{2\mu_s}, \qquad \theta_{\min} = \arctan\frac{1}{2(0.40)} = \arctan 1.25 \approx \boxed{51^\circ}. \tag{12.1.13}$$

Below this angle the ladder slides; the result is independent of the ladder's mass and length.

Check. A shallower ladder has a larger $\cos\theta$ and smaller $\sin\theta$, so the weight's lever arm grows while the wall's shrinks; $F_{N,w}$, and hence the friction needed, rises. That is why ladders slip when leaned too far. With a perfectly slippery floor ($\mu_s \to 0$), $\tan\theta \to \infty$ and $\theta_{\min} \to 90^\circ$: the ladder would have to stand vertical.

Example 12.4: A strut and a rope

A uniform strut of weight $780\ \text{N}$ is pinned to a wall at its lower end and rises at 60° above the horizontal. A load $F_{g,\text{load}} = 815\ \text{N}$ hangs from its upper end, and a rope attached at that same end (perpendicular to the strut) runs back to the wall. Find the tension in the rope.



•Figure description
A strut projecting from a wall at 60 degrees with its own weight F sub g,strut acting at its midpoint and a load of weight F sub g,load hanging from its far end; the 60 degree angle is measured from a dashed horizontal line through the pin. A rope attached at the far end runs perpendicular to the strut back to the wall, under tension F sub T.

Solution

Step 1: Choose the pivot. Take torques about the pin P, eliminating the unknown wall force.

Step 2: Lever arms. The rope is perpendicular to the strut, so its lever arm is the full length L: its torque is $F_T L$. The strut's weight acts at its center (distance $L/2$ along the strut) and the load at the end (distance L). Both of these forces are vertical, so their lever arms about P are not the distances along the strut but the *horizontal* distances from P to their lines of action. The strut rises at 60°, so the horizontal distance to a point a distance s along it is $s\cos 60^\circ$ (the side adjacent to the angle): $\frac{L}{2}\cos 60^\circ$ and $L\cos 60^\circ$.

Step 3: Balance torques. The rope's torque balances the torques of the two weights:

$$F_T L = F_{g,\text{load}}\,(L\cos 60^\circ) + F_{g,\text{strut}}\left(\tfrac{L}{2}\cos 60^\circ\right). \tag{12.1.14}$$

Every term contains L, so divide it out: the answer cannot depend on the strut's length. Then factor $\cos 60^\circ$ out of the right side:

$$\begin{aligned} F_T &= \left(F_{g,\text{load}} + \tfrac{1}{2}F_{g,\text{strut}}\right)\cos 60^\circ = \left(815 + \tfrac{1}{2}(780)\right)(0.500)\,\text{N} \\ &= (815+390)(0.500)\,\text{N} \approx \boxed{603\,\text{N}}. \end{aligned} \tag{12.1.15}$$

The strut's weight counts only half as much as the load because it acts, in effect, halfway along the strut.

Answers to Checkpoints

Checkpoint 12.1. Static equilibrium requires zero net *torque*; moment of inertia only matters when the system has an angular acceleration, which here it does not. Each cord pulls tangentially at its disk's rim, so the two torques are $mg\ r$ and $Mg\ R$, and balance requires $mgr = MgR$, that is, $mr = MR$. The lever arm enters to the first power, not squared: the heavier block must hang from the smaller disk in inverse proportion to the radii.

Checkpoint 12.2. Take torques about the pin. The pin force acts *at* the pin, so whatever its size and direction, its lever arm about that point is zero and its torque vanishes exactly. It does not appear in the torque equation at all, so the only unknown left is the cable tension. This is the general strategy: choose the pivot at the point where the most unknown forces act.

This page titled 12.1: Equilibrium is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12.2: Center of Gravity and Stability

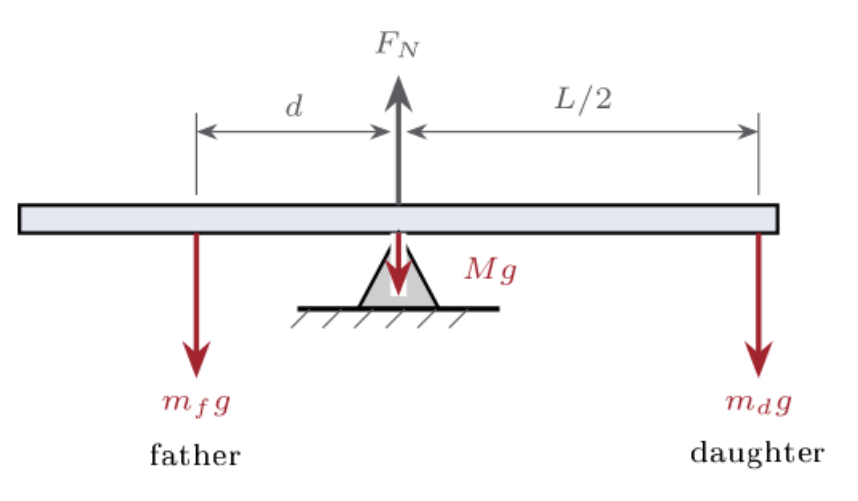
Throughout the last section we treated each object's weight as though it acted at a single point. That point is the **center of gravity** (cg): the point where the total gravitational force can be considered to act. Its position is the weight-weighted average of the positions of all the parts:

$$x_{\text{cg}} = \frac{\sum m_i g\, x_i}{\sum m_i g}. \tag{12.2}$$

As long as g is the same everywhere on the object (true for any everyday object), the g's cancel and the center of gravity coincides exactly with the **center of mass** of §9.5. This is why we can replace an extended body's distributed weight with a single force mg at its center of mass when computing torques. For a uniform, symmetric object the cg sits at the geometric center; for a composite object it is found by treating each part's weight as acting at that part's own center.

Example 12.5: A father-daughter seesaw

A seesaw is a uniform board of mass M and length L, pivoted at its center (its center of gravity). A father of mass m_f sits a distance d from the center. His daughter, of mass m_d, sits on the *opposite* side, at the very end of the board: a distance $L/2$ from the center. (a) Where must the father sit for balance? (b) What upward force does the pivot supply?



•Figure description
A seesaw board of length L balanced on a pivot at its center. The father's weight m sub f g acts downward a distance d to the left of the pivot; the daughter's weight m sub d g acts downward at the right end, a distance L over 2 from the pivot. The board's own weight M g acts at the pivot, and the pivot pushes up with force F sub N.

Solution

(a) Because the board is pivoted at its own center of gravity, the board's weight produces no torque about the pivot (its lever arm is zero). Balancing the father's and daughter's torques:

$$m_f\, g\, d = m_d\, g\, \frac{L}{2}. \tag{12.2.1}$$

Divide both sides by $m_f g$; the g cancels:

$$d = \boxed{\frac{m_d L}{2 m_f}}. \tag{12.2.2}$$

A heavier father sits closer to the pivot, as experience confirms. The units check: masses divide out, leaving a length. And if the

two masses were equal, $d = L/2$: the father would sit at the other end, as symmetry demands.

(b) The pivot must support the entire weight, since the net vertical force is zero: $F_N - Mg - m_f g - m_d g = 0$, so

$$F_N = \boxed{(M + m_f + m_d)\, g}. \tag{12.2.3}$$

Note that this does not depend on where anyone sits: moving the riders around redistributes torque, not weight.

Stability and the base of support

Knowing where the center of gravity sits does more than let us compute torques: it tells us whether an object will stay put at all. An object resting on a surface touches it over some region, and the **base of support** is the smallest area enclosing all of those contact points. For a filing cabinet it is the rectangle under the cabinet; for a person standing it is the region spanned by both feet; for a tripod it is the triangle joining the three feet.

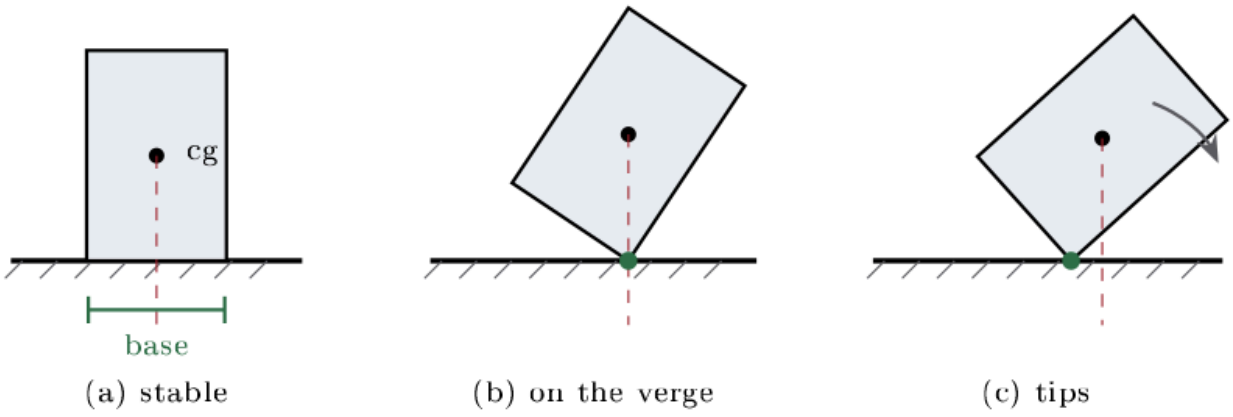


Figure 12.1. Tipping. The dashed line is the vertical through the center of gravity. In (a) it lands inside the base of support and the block stands. In (b) the block has been tilted just far enough that the line passes through the contact edge: the critical angle. In (c) the line has fallen outside the base, and gravity's torque about the edge rotates the block over. Tilting the block is equivalent to placing it on a ramp of the same angle.

•Figure description

Two panels on tipping. In panel (a) the dashed vertical line through the center of gravity falls inside the base of support and the block is stable. In panel (b) the block is tilted further, the line falls outside the base, and the block topples.

Now picture the vertical line running downward through the center of gravity. The weight acts along that line, and the supporting normal forces act somewhere within the base. So long as the line falls *inside* the base, the support can distribute itself to balance the weight with no leftover torque, and the object stands. Tilt the object far enough that the line passes *outside* the base, and it cannot: the contact has shifted to a single edge, and gravity's torque about that edge is unopposed. The object tips (Figure 12.1).

For a uniform rectangular block of width w and height h standing on a surface that is slowly tilted, the center of gravity is a horizontal distance $w/2$ from the downhill edge and a height $h/2$ above the base. The vertical through it reaches that edge when

Key Equation: Critical tipping angle for a block

$$\tan\theta_c = \frac{w}{h}. \tag{12.3}$$

Wide and low means a large θ_c and a stable object; tall and narrow means a small one. This single result explains a great deal of practical design: why sports cars are built low and wide, why a loaded moving truck is far more prone to rollover than an empty one (loading raises h), why a stepladder splays its feet outward, and why you widen your stance on a moving bus.

It is also worth naming the three ways an object can sit in equilibrium. If a small displacement *raises* the center of gravity, gravity pushes the object back and the equilibrium is **stable**: a marble in a bowl, or the block of Figure 12.1(a). If the displacement *lowers* it, the object keeps going and the equilibrium is **unstable**: a pencil balanced on its tip, or the block in (c). If the height

is unchanged, the equilibrium is **neutral**: a ball on a level table is content anywhere. Case (b), exactly at the critical angle, is the unstable boundary between the first two.

Example 12.6: Will the crate slide or tip?

A uniform crate $0.80\ \text{m}$ wide and $1.6\ \text{m}$ tall sits on a ramp whose angle is slowly increased from zero. The coefficient of static friction between crate and ramp is $\mu_s = 0.60$. Which happens first, sliding or tipping, and at what angle?

Solution

Treat the two failures separately and see which one the ramp angle reaches first.

Sliding. On an incline the crate begins to slide when the component of gravity along the ramp overcomes maximum static friction, which (from Chapter 6) happens at

$$\tan\theta_{\text{slide}} = \mu_s = 0.60 \implies \theta_{\text{slide}} = \arctan 0.60 = 31.0^\circ. \tag{12.2.4}$$

Tipping. From equation (12.3), the vertical through the center of gravity leaves the base at

$$\tan\theta_{\text{c}} = \frac{w}{h} = \frac{0.80\,\text{m}}{1.6\,\text{m}} = 0.50 \implies \theta_{\text{c}} = \arctan 0.50 = 26.6^\circ. \tag{12.2.5}$$

Tipping happens at the smaller angle, so the crate $\boxed{\text{tips first, at } 26.6^\circ}$; it never gets the chance to slide.

A rule from the comparison. Between 0° and 90° the tangent always increases with the angle, so comparing two angles is the same as comparing their tangents. The crate tips first exactly when $\tan\theta_{\text{c}} < \tan\theta_{\text{slide}}$, that is, whenever $w/h < \mu_s$. There was no need to compute either angle to answer "which first": $0.50 < 0.60$ settles it. Tall, narrow objects on grippy surfaces tip; short, wide objects on slippery ones slide.

Checkpoint 12.3

A hiker puts on a heavy backpack and immediately leans forward; a waiter carrying a loaded tray on one side leans away from it. Neither is thinking about physics. Using the base of support, explain what both are actually accomplishing, and explain why a person carrying a heavy suitcase in one hand cannot simply stand up straight.

Answer

Answers to Checkpoints

Checkpoint 12.3. Both are keeping the combined center of gravity (their body plus the load) directly above the base of support, their feet. The backpack moves the combined center of gravity backward, so the hiker leans forward to bring it back over the feet; the waiter leans away from the tray for the same reason. A person holding a heavy suitcase in one hand who stood perfectly straight would have the combined center of gravity shifted toward the suitcase, near or beyond the edge of the feet. Gravity would then exert a torque about that edge and tip the person toward the suitcase, so the person must lean the other way.

This page titled 12.2: Center of Gravity and Stability is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12.3: Elasticity

No real object is perfectly rigid. Apply forces to a solid and it deforms: a little or a lot, reversibly or permanently. **Elasticity** is the study of how solids change shape under load and spring back when the load is removed. The central idea is a proportionality, within limits, between the **stress** applied (force per unit area) and the resulting **strain** (fractional deformation). Their ratio is a property of the material called an **elastic modulus**:

$$\text{elastic modulus} = \frac{\text{stress}}{\text{strain}}. \tag{12.3.1}$$

This is Hooke's law generalized from springs to materials: double the stress and you double the strain, until the *elastic limit* is reached and the material yields or breaks.

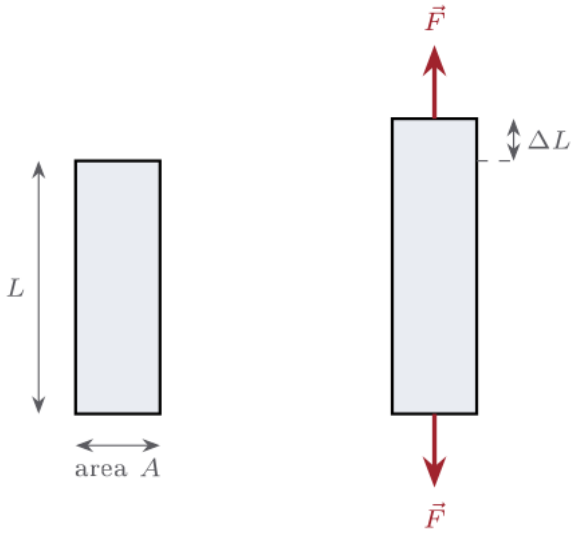


Figure 12.2. Tensile stress and strain. A force F applied to a bar of cross-sectional area A and length L stretches it by ΔL. Young's modulus is the ratio of the stress F/A to the strain $\Delta L/L$.

•Figure description

A bar of length L and cross-sectional area A pulled by equal and opposite forces F at each end, stretching it by an amount delta L.

There are three common deformations, each with its own modulus. For stretching or compressing along a length (Figure 12.2), **Young's modulus** Y relates the tensile stress F/A to the fractional change in length $\Delta L/L$:

Key Equation: Young's modulus

$$Y = \frac{F/A}{\Delta L/L}. \tag{12.4}$$

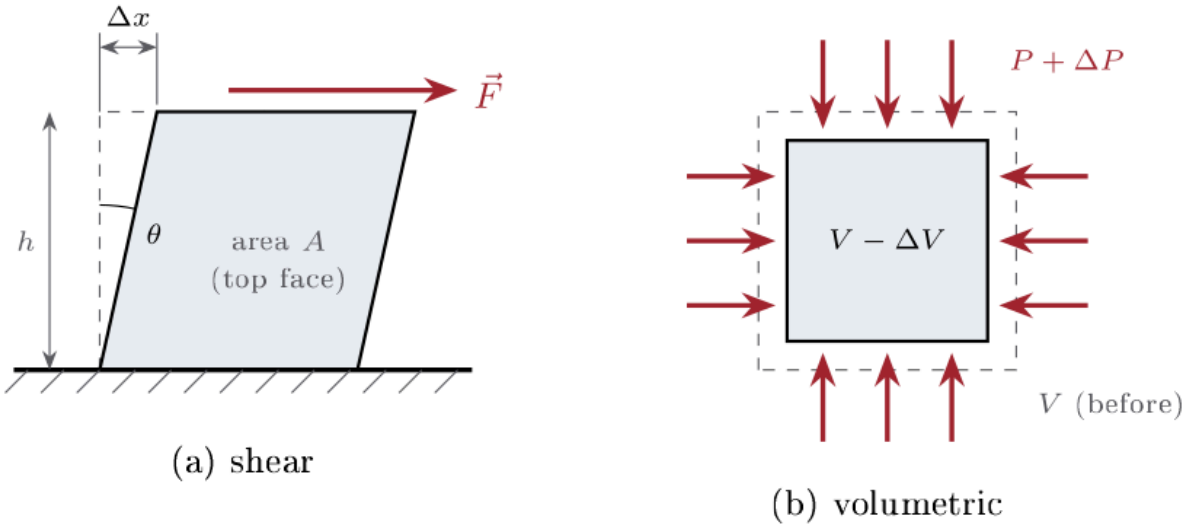


Figure 12.3. The other two deformations. (a) In *shear*, the force acts parallel to a face of area A, sliding it a distance Δx relative to a face h away; the dashed outline is the undeformed shape, and the strain $\Delta x/h$ equals $\tan\theta$. (b) Under a uniform pressure increase ΔP the object shrinks from volume V to $V - \Delta V$ without changing shape. Both deformations are drawn greatly exaggerated.

•Figure description

Two panels showing shear and volume deformation. In shear, a force parallel to a face of area A slides it a distance delta x relative to the opposite face across height h. In volume deformation, pressure acts inward on all faces, reducing the volume.

Stress has units of pascals ($\mathrm{Pa} = \mathrm{N/m^2}$), and strain is dimensionless, so the moduli are measured in pascals. Young's modulus is large for stiff materials (about 2×10^{11} Pa for steel), meaning enormous stress is needed for even a slight stretch. The other two cases follow the same pattern. The **shear modulus** S governs a sideways deformation. Here the force F is applied *parallel* to a face of area A rather than perpendicular to it, sliding that face a distance Δx relative to a parallel face a distance h away, so that the object's square cross-section becomes a parallelogram (Figure 12.3a):

Key Equation: Shear modulus

$$S = \frac{F/A}{\Delta x/h}. \tag{12.5}$$

Note that A here is the area of the face the force acts *along*, not the area it acts across. The shear strain $\Delta x/h$ is the tangent of the small angle through which the sides tilt, so for the tiny deformations of stiff materials it is simply that angle in radians.

And the **bulk modulus** B governs uniform compression, in which pressure is applied over the entire surface and the object shrinks without changing shape (Figure 12.3b). It relates the change in pressure ΔP to the fractional change in volume:

Key Equation: Bulk modulus

$$B = -\frac{\Delta P}{\Delta V/V}. \tag{12.6}$$

The minus sign makes B positive, since an increase in pressure decreases the volume. A liquid or gas has a bulk modulus but no Young's or shear modulus: a fluid at rest cannot sustain a tension or a static shear, which is exactly what makes it flow. Solids have all three.

Table 12.1 collects representative values. Two patterns are worth noticing. First, for a given solid the three moduli are of the same order of magnitude, with S typically a third to a half of Y. Second, the range across materials is enormous (steel is about seventy times stiffer than nylon), which is why material choice, not just geometry, decides how much a structure deflects.

Table 12.1. Approximate elastic moduli for selected materials. Values vary with composition, temperature, and processing, so treat these as representative rather than exact. Liquids have no Young's or shear modulus because they cannot sustain a static tension or shear. Concrete's Young's modulus is quoted in compression; in tension it is far weaker and cracks at low stress, which is why structural concrete is reinforced with steel.

Material	Y (Pa)	S (Pa)	B (Pa)
Tungsten	3.5×10^{11}	1.4×10^{11}	2.0×10^{11}
Steel	2.0×10^{11}	8.4×10^{10}	1.6×10^{11}
Copper	1.1×10^{11}	4.4×10^{10}	1.4×10^{11}
Brass	9.0×10^{10}	3.5×10^{10}	6.0×10^{10}
Aluminum	7.0×10^{10}	2.5×10^{10}	7.0×10^{10}
Glass	6.5×10^{10}	2.6×10^{10}	5.0×10^{10}
Concrete	2.0×10^{10}	—	—
Bone (compact)	1.6×10^{10}	—	—
Nylon	3.0×10^{9}	—	—
Water	—	—	2.2×10^{9}
Mercury	—	—	2.8×10^{10}

Finally, notice what Young's modulus says about an *object* rather than a material. Solving equation (12.4) for the force gives

$$F = \left(\frac{YA}{L}\right)\Delta L, \tag{12.3.2}$$

which is exactly Hooke's law $F = k\ \Delta L$ with an effective spring constant

$$k = \frac{YA}{L}. \tag{12.7}$$

Every rod, wire, and cable is a stiff spring. The modulus Y belongs to the material; the stiffness k belongs to the particular piece, and grows with its cross-section and shrinks with its length. This is why a long cable sags more than a short one of the same steel, and it lets us carry over everything we know about springs, including the rule that springs joined end to end add as $1/k = 1/k_1 + 1/k_2$.

Checkpoint 12.4

An engineer replaces a support wire with one of the same material and length but twice the diameter, carrying the same load. By what factor does each of the following change: the stress in the wire, the strain, the elongation, and the effective spring constant k? Which of these four quantities, if any, is a property of the material rather than of this particular wire?

Answer

Example 12.7: Stretch of a steel rod

A steel support rod $2.0\ \text{m}$ long with a cross-sectional area of $3.0 \times 10^{-4}\ \text{m}^2$ carries a $1500\ \text{kg}$ load. Taking $Y_{\text{steel}} = 2.0 \times 10^{11}\ \text{Pa}$, find how much it stretches.

Solution

Step 1: The force. The rod supports the load, so the tensile force equals the load's weight: $F = mg = (1500\ \text{kg})(9.8\ \text{m/s}^2) = 14700\ \text{N}$.

Step 2: Rearrange Young's modulus. Equation (12.4) defines Y as stress over strain, $Y = (F/A)/(\Delta L/L)$. Dividing by a fraction is multiplying by its reciprocal, so this is $Y = FL/(A\ \Delta L)$. To isolate ΔL, multiply both sides by ΔL and divide both sides by Y:

$$Y\,\Delta L = \frac{FL}{A} \implies \Delta L = \frac{FL}{AY}. \tag{12.3.3}$$

Step 3: Substitute. Work out the numerator and denominator separately, collecting powers of ten in the denominator ($10^{-4} \times 10^{11} = 10^{7}$):

$$\Delta L = \frac{(14700\,\mathrm{N})(2.0\,\mathrm{m})}{(3.0\times10^{-4}\,\mathrm{m^2})(2.0\times10^{11}\,\mathrm{Pa})} = \frac{29400\,\mathrm{N\cdot m}}{6.0\times10^{7}\,\mathrm{N}} \approx 4.9\times10^{-4}\,\mathrm{m} \approx \boxed{0.49\,\mathrm{mm}}. \tag{12.3.4}$$

(A pascal is $\mathrm{N/m^2}$, so $\mathrm{m^2\cdot\ Pa} = \mathrm{N}$, and the units reduce to meters.)

Half a millimeter under a load of $1500\ \mathrm{kg}$: a vivid reminder of how stiff steel is, and why it carries our buildings and bridges. It is worth checking that we are entitled to use equation (12.4) at all: the stress here is $F/A = 14700/(3.0\times10^{-4}) = 4.9\times10^{7}$ Pa, roughly a fifth of the stress at which structural steel begins to yield, so the rod is comfortably in its elastic range and the linear relation holds.

Beyond the elastic limit

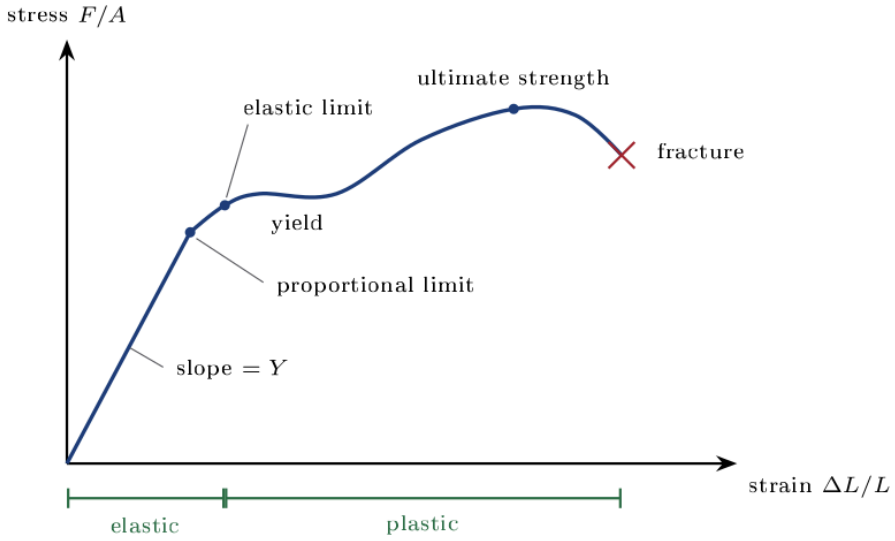


Figure 12.4. Stress versus strain for a ductile metal loaded to fracture. The initial straight portion is the elastic region of equation (12.4), and its slope is Young's modulus. Past the elastic limit the deformation is partly permanent; the material then yields, strain hardens to its ultimate strength, necks down, and breaks. The strain axis is greatly compressed: for steel the elastic region occupies well under one percent strain, while the plastic region can run to twenty percent or more.

•Figure description

A stress-strain curve for a ductile metal loaded to fracture. It rises as a straight line of slope Y through the elastic region to the proportional limit and elastic limit, then bends over at the yield point, continues to a maximum at the ultimate strength, and falls to fracture.

Everything above has assumed that stress and strain stay proportional. Figure 12.4 shows what actually happens when a ductile metal (mild steel, say) is pulled steadily until it breaks. The linear behavior we have been using is only the first stretch of the story.

The curve starts as a straight line whose slope is Y, and stays straight up to the **proportional limit**. Beyond that point stress and strain are no longer proportional and equation (12.4) no longer applies. Slightly further along lies the **elastic limit**. Up to there the specimen still springs back to its original length when the load is removed; past it, it does not. Unload a specimen that has been stretched into that region and it relaxes along a line *parallel* to the original straight portion, coming to rest at a nonzero strain: a **permanent set**. This is the difference between bending a paper clip slightly, which does nothing, and bending it firmly, which leaves it bent.

Just past the elastic limit a ductile metal **yields**: it deforms a great deal with almost no increase in stress, producing the flat stretch of the curve. The stress at which this begins, the **yield strength**, is usually the number a designer cares about, since a beam that has yielded is permanently deformed even though it has not broken. Beyond the yield region the metal stiffens again (*strain hardening*) and the stress climbs to a maximum, the **ultimate strength**. After that the specimen thins locally, or *necks*, and the curve falls until it fractures.

Not all materials behave this way. Glass, cast iron, and concrete loaded in tension are **brittle**: they have almost no plastic region and fracture at or near the elastic limit, with little warning. Ductile materials announce trouble by deforming visibly first, which is one reason structural steel rather than cast iron carries our bridges.

Engineers do not design a member to sit anywhere near these limits. The yield strength (or sometimes the ultimate strength) is divided by a **safety factor** (commonly between 2 and 10, depending on how well the loads are known, how variable the material is, and how serious a failure would be) to give an allowable working stress, and the member is sized so that the actual stress stays below it. The rod of Example 12.7 carries about a fifth of its yield stress, a safety factor of roughly 5.

Checkpoint 12.5

A wire is stretched to a strain just past its elastic limit and then completely unloaded. Sketch the loading and unloading paths on a stress–strain diagram. Is the wire now longer than it started? If you reload it, at roughly what stress does it begin to deform plastically again: the original yield strength, or a higher one? What does your answer say about why cold-working a metal makes it harder?

Answer

Answers to Checkpoints

Checkpoint 12.4. Doubling the diameter quadruples the cross-sectional area, $A = \pi d^2/4$. With the same load, the stress F/A falls to $\frac{1}{4}$; the strain, which is stress divided by Y, also falls to $\frac{1}{4}$; the elongation, which is strain times the same length, falls to $\frac{1}{4}$; and the effective spring constant $k = YA/L$ increases by a factor of 4. None of the four is a property of the material alone: each depends on the load or on the wire's dimensions. The material property is Young's modulus Y, which is the same for both wires.

Checkpoint 12.5. Loading follows the straight elastic line up to the elastic limit, then bends over as the wire yields. On unloading, the stress falls along a straight line *parallel* to the original elastic line, reaching zero stress at a positive strain. So yes, the wire is permanently longer than it started. On reloading, it follows that same parallel line back up and behaves elastically until the stress reaches the highest value it saw before, which is above the original yield strength. The wire has become harder to deform permanently. Cold-working a metal (rolling, drawing, hammering) does exactly this: each plastic deformation raises the stress needed to cause the next one.

This page titled 12.3: Elasticity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12.S: Static Equilibrium and Elasticity (Summary)

Summary

- An object is in **static equilibrium** when $\sum \mathbf{F} = 0$ *and* $\sum \tau = 0$. Because the net torque vanishes about every point, choosing the axis at an unknown force eliminates it from the torque equation.
- In two dimensions these give exactly three independent equations, so statics alone can determine at most three unknowns. Problems with more are **statically indeterminate** and need elasticity as well.
- The **center of gravity** is where an object's weight effectively acts; for uniform g it coincides with the center of mass. We replace distributed weight with a single force mg there.
- An object **tips** when the vertical line through its center of gravity falls outside its **base of support**. For a block of width w and height h this happens at $\tan\theta_c = w/h$ (12.3); low and wide is stable.
- **Stress** (force per area) and **strain** (fractional deformation) are proportional within the elastic limit, with ratio the **elastic modulus**.
- **Young's modulus** $Y = (F/A)/(\Delta L/L)$ (12.4) governs stretching; the **shear modulus** $S = (F/A)/(\Delta x/h)$ (12.5) governs sideways deformation; and the **bulk modulus** $B = -\Delta P/(\Delta V/V)$ (12.6) governs volume compression.
- Any rod or wire behaves as a spring of stiffness $k = YA/L$ (12.7): the modulus is a property of the material, the stiffness a property of the piece.
- The linear relation holds only up to the **proportional limit**. Past the **elastic limit** deformation is permanent; a ductile material then **yields**, strain hardens to its **ultimate strength**, and fractures. Designers work to a fraction of the yield strength set by a **safety factor**.

Key Equations

Conditions for static equilibrium

$$\sum \mathbf{F} = 0 \qquad \text{and} \qquad \sum \tau = 0 \tag{12.S.1}$$

Critical tipping angle for a block

$$\tan\theta_c = \frac{w}{h} \tag{12.S.2}$$

Young's modulus

$$Y = \frac{F/A}{\Delta L/L} \tag{12.S.3}$$

Shear modulus

$$S = \frac{F/A}{\Delta x/h} \tag{12.S.4}$$

Bulk modulus

$$B = -\frac{\Delta P}{\Delta V/V} \tag{12.S.5}$$

This page titled 12.S: Static Equilibrium and Elasticity (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

12.E: Static Equilibrium and Elasticity (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 12.1 — Equilibrium

12.1 ★ *[Conceptual.]* State the two conditions for static equilibrium. Give an everyday example of an object that satisfies the force condition but *not* the torque condition, and describe what it does.

12.2 ★ A $3.0\ \text{m}$ uniform plank of weight $120\ \text{N}$ rests on two supports, one at each end. A $200\ \text{N}$ child stands $1.0\ \text{m}$ from the left support. Find the upward force from each support.

12.3 ★★ A uniform beam of mass $15\ \text{kg}$ and length $4.0\ \text{m}$ is hinged at a wall and held horizontal by a vertical cable at its far end.

a. Find the tension in the cable.
b. Find the vertical force the hinge exerts on the beam.

12.4 ★★ *[Real-world: construction.]* A $500\ \text{N}$ uniform horizontal beam $6.0\ \text{m}$ long is pinned to a wall, with its far end held by a cable at 40° above the beam. A $700\ \text{N}$ load hangs from the far end. Find the cable tension.

12.5 ★★ A uniform ladder leans against a frictionless wall at 60° above the ground. Find the minimum coefficient of static friction with the ground that keeps it from slipping.

12.6 ★★★ A $1200\ \text{N}$ uniform beam $5.0\ \text{m}$ long is supported by a cable at its far end making 30° with the beam; the near end is pinned to a wall. A $2000\ \text{N}$ load sits $3.5\ \text{m}$ from the wall.

a. Find the cable tension.
b. Find the magnitude and direction of the force at the pin.

12.7 ★★ *[Conceptual.]* When you carry a heavy bucket in one hand, you instinctively lean your body the other way. Explain this in terms of keeping your center of gravity over your feet and the conditions for equilibrium.

12.8 ★★★ *[Challenge.]* A uniform beam of mass m and length L is held horizontal by a pivot a distance $L/4$ from its left end. A support line attached to the beam's right end runs upward and to the left at an angle θ above the horizontal, and a force sensor reads the line's tension.

a. Sketch the beam and identify the lever arm of the beam's weight about the pivot. Explain why a uniform beam has a nonzero weight torque here, even though it had none in the seesaw of Example 12.5.
b. Find the tension in the support line in terms of m, g, and θ.
c. Find the x- and y-components of the force the pivot exerts on the beam.

Section 12.2 — Center of Gravity and Stability

12.9 ★ A uniform meter stick is balanced at its center. A $50\ \text{g}$ mass is hung at the $20\ \text{cm}$ mark. Where must a $100\ \text{g}$ mass be hung to rebalance it?

12.10 ★ *[Conceptual.]* Why does a car with a low center of gravity corner more safely than one with a high center of gravity? Relate your answer to the tendency to tip (rotate) versus skid.

12.11 ★★ Two masses, $2.0\ \text{kg}$ and $6.0\ \text{kg}$, are attached to the ends of a light rod $1.2\ \text{m}$ long. Find the location of the center of gravity (measured from the lighter mass), and explain why a single support there would balance the rod.

12.12 ★★ *[Real-world: sports.]* A high jumper arches her body into a curve as she crosses the bar. Explain how this lets her center of gravity pass *below* the bar even as her body passes over it, and why that is an advantage.

12.13 ★★ A uniform beam of mass $8.0\ \text{kg}$ and length $2.0\ \text{m}$ has a $4.0\ \text{kg}$ point mass fixed to one end. Find the center of gravity of the combination, measured from the bare end.

Section 12.3 — Elasticity

12.14 ★ A wire $1.5\ \text{m}$ long and of cross-sectional area $2.0 \times 10^{-6}\ \text{m}^2$ stretches by $1.0\ \text{mm}$ under a $250\ \text{N}$ load. Find Young's modulus for the wire's material.

12.15 ★ *[Conceptual.]* Define stress and strain, and explain why the elastic modulus (their ratio) is a property of the material rather than of the particular object.

12.16 ★★ A steel cable ($Y = 2.0 \times 10^{11}\ \text{Pa}$) of diameter $1.0\ \text{cm}$ and length $20\ \text{m}$ supports a $2500\ \text{kg}$ elevator car.

a. Find the stress in the cable.
b. Find how much the cable stretches.

(Compare your answer to (a) with the roughly $2.5 \times 10^{8}\ \text{Pa}$ at which structural steel begins to yield. This is why a real elevator hangs from several cables rather than one.)

12.17 ★★ *[Real-world: materials.]* A nylon guitar string ($Y = 3.0 \times 10^{9}\ \text{Pa}$) and a steel one ($Y = 2.0 \times 10^{11}\ \text{Pa}$) have the same length and cross-section and carry the same tension. Which stretches more, and by what factor?

12.18 ★★ The bulk modulus of water is $2.2 \times 10^{9}\ \text{Pa}$. Find the fractional change in the volume of a sample of water taken to a depth where the pressure is $4.0 \times 10^{7}\ \text{Pa}$ greater than at the surface.

12.19 ★★★ A solid brass cube ($S = 3.5 \times 10^{10}\ \text{Pa}$) $6.0\ \text{cm}$ on a side has its top face pushed sideways by a force of $2.0 \times 10^{4}\ \text{N}$ parallel to the face while the bottom is held fixed. Find the horizontal displacement of the top face.

12.20 ★ *[Conceptual.]* Two rods are made of the same material and have the same length, but one has twice the cross-sectional area of the other. Under the same applied force, which rod stretches more, and by what factor? Identify which quantity (stress or strain) differs between them, and which is the same.

Additional Problems

12.21 ★★★ A uniform ladder of length $5.0\ \text{m}$ and weight $150\ \text{N}$ leans at 60° above the floor against a frictionless wall. The coefficient of static friction with the floor is 0.40. A $700\ \text{N}$ person climbs the ladder.

a. Write the equilibrium conditions, identifying the floor normal force, the floor friction, and the wall normal force.
b. How far up the ladder (measured along it) can the person climb before it begins to slip?
c. What fraction of the ladder's length is that?
d. What minimum floor friction coefficient would let the person reach the very top?

12.22 ★★★ *[Real-world: engineering.]* A $1500\ \text{kg}$ car has a wheelbase (front-to-rear axle distance) of $2.6\ \text{m}$. A scale under the front axle reads $9000\ \text{N}$ and one under the rear axle reads $5700\ \text{N}$.

a. Verify that the readings are consistent with the car's weight.
b. How far behind the front axle is the center of gravity?
c. What fraction of the weight does each axle carry?
d. If $300\ \text{kg}$ of cargo is placed directly over the rear axle, where is the new center of gravity?

12.23 ★★★ *[Real-world: materials.]* A $200\ \text{N}$ load hangs from a steel wire ($Y = 2.0 \times 10^{11}\ \text{Pa}$, length $3.0\ \text{m}$) joined end to end to a brass wire ($Y = 9.0 \times 10^{10}\ \text{Pa}$, length $2.0\ \text{m}$); both have cross-sectional area $2.0 \times 10^{-6}\ \text{m}^2$.

a. Find the stress in each wire, and explain why the two are equal.
b. Find the elongation of each wire.
c. Find the total elongation.
d. Using equation (12.7), find the effective spring constant of the combination, and verify that it satisfies the series rule $1/k = 1/k_1 + 1/k_2$.

12.24 ★★ *[Real-world: safety.]* A uniform filing cabinet $0.45\ \text{m}$ wide and $1.5\ \text{m}$ tall stands on a level floor.

a. Through what angle must the floor be tilted before the cabinet tips over?
b. Pulling out a loaded top drawer shifts the cabinet's center of gravity $0.10\ \text{m}$ horizontally toward the tipping edge and $0.20\ \text{m}$ upward. Find the new tipping angle.
c. Your answers should explain a common safety warning about tall cabinets. State it, and say why anchoring the cabinet to the wall works.

12.25 ★★★ *[Real-world: vehicle dynamics.]* A delivery truck has a track width (left wheels to right wheels) of $t = 1.9\ \text{m}$ and a center of gravity $h = 1.1\ \text{m}$ above the road. It rounds a flat curve, so friction supplies a horizontal acceleration a toward the center.

a. Let $F_{N,\text{in}}$ and $F_{N,\text{out}}$ be the normal forces on the inner and outer wheels. Writing $\sum \tau = 0$ about the center of gravity and $\sum F = 0$, show that the truck is on the verge of rolling over (that is, $F_{N,\text{in}} = 0$) when $a = gt/2h$.
b. Evaluate this rollover threshold for the truck, as a multiple of g.
c. The truck instead skids when a exceeds $\mu_s g$. For $\mu_s = 0.85$, does it roll or skid first?
d. A load raises the center of gravity to $1.5\ \text{m}$. Repeat (b) and (c), and comment on what this means for how cargo should be stacked.

12.26 ★★ *[Real-world: design.]* A steel tie rod must carry a static load of $4.0 \times 10^{4}\ \text{N}$. The steel has a yield strength of $2.5 \times 10^{8}\ \text{Pa}$, and the design code calls for a safety factor of 3 against yielding.

a. Find the maximum allowable working stress.
b. Find the minimum cross-sectional area, and the diameter of the smallest round rod that will do.
c. The rod is $2.5\ \text{m}$ long. How much does it stretch at the working load? ($Y = 2.0 \times 10^{11}\ \text{Pa}$.)

This page titled 12.E: Static Equilibrium and Elasticity (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13: UNIVERSAL GRAVITATION

Gravity is the most familiar of forces (it holds us to the ground from the moment we are born), yet it took a leap of imagination to see what it really is. Newton's insight was not that gravity *exists*, but that it is **universal**: the same force that pulls an apple to the ground reaches out across empty space to hold the Moon in its orbit, to bind the planets to the Sun, and to gather stars into galaxies. The Moon, Newton realized, is perpetually *falling* toward the Earth (falling below the straight-line path it would otherwise follow), but its sideways motion carries it around instead of into us.

This chapter develops that single law and follows its consequences. We will compute the gravitational force between any two masses, use it to find the free-fall acceleration on any world and the speeds of satellites, account for the energy of orbiting bodies, and finally arrive at Kepler's three laws of planetary motion: empirical rules that fall out of Newton's gravity like theorems from an axiom.

Learning Objectives

- State and apply Newton's law of universal gravitation, including the vector sum of several gravitational forces.
- Explain why a spherical planet may be treated as a point mass at its center.
- Calculate the free-fall acceleration on a planet or moon and the orbital speed of a satellite.
- Use gravitational potential energy and energy conservation for planets, satellites, and escape.
- State Kepler's three laws and use the third to relate orbital periods, radii, and masses.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

13.1: Newton's Law of Universal Gravitation
13.2: Applications of Universal Gravitation
13.3: Gravitational Potential Energy
13.4: Kepler's Laws of Planetary Motion
13.S: Universal Gravitation (Summary)
13.E: Universal Gravitation (Problems)

This page titled 13: Universal Gravitation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.1: Newton's Law of Universal Gravitation

Every body in the universe attracts every other body. For two particles of masses m_1 and m_2 separated by a distance r, the attraction is proportional to the product of the masses and inversely proportional to the square of the distance between them:

Key Equation: Newton's law of universal gravitation

$$F_g = G\frac{m_1 m_2}{r^2}. \tag{13.1}$$

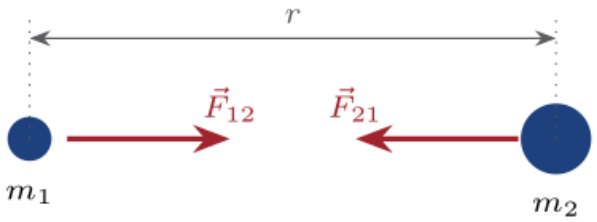


Figure 13.1. Two masses attract each other with equal and opposite forces directed along the line joining them, of magnitude Gm_1m_2/r^2. Note that r is measured *center to center*, not surface to surface.

•Figure description

Two masses m sub 1 and m sub 2 separated by a distance r, each pulled toward the other by an equal and opposite force directed along the line joining them.

The force on each body points toward the other (gravity is always attractive) and the two forces form a Newton's-third-law pair, equal in magnitude and opposite in direction (Figure 13.1). We write $\mathbf{F}_{12}$ for the force *on* body 1 *due to* body 2: the first subscript names the body being pushed on, the second names the body doing the pulling. In that notation Newton's third law reads $\mathbf{F}_{12} = -\mathbf{F}_{21}$. Keeping the subscripts in that order will save you a great deal of grief once three or more masses are in play.

The constant of proportionality is the **universal gravitational constant**,

$$G = 6.674 \times 10^{-11}\,\mathrm{N\cdot m^2/kg^2}, \tag{13.1.1}$$

a tiny number, which is why gravity is utterly negligible between everyday objects and becomes commanding only when at least one mass is astronomical.

The $1/r^2$ dependence is the **inverse-square law**, and it has a simple geometric origin: an influence spreading out from a point source is diluted over a spherical surface whose area grows as r^2. Doubling the distance spreads the same gravity over four times the area, weakening it to a quarter. The force shrinks rapidly with distance but never quite reaches zero: gravity has, in principle, infinite range.

Note: Why we may treat a planet as a point

Equation (13.1) is written for *particles*, yet we will apply it to planets, moons, and billiard balls, always measuring r from center to center. The justification is a theorem Newton proved for exactly this purpose: a body with spherically symmetric density attracts anything *outside* it precisely as though all of its mass were concentrated at its center. The near side pulls harder and the far side pulls softer, and the geometry makes the two effects cancel exactly.

This is why $g = Gm_1/R^2$ works at a planet's surface even though most of the planet lies thousands of kilometers away, and it is the only reason the calculations in this chapter are as simple as they are. Newton also proved the companion result: *inside* a uniform spherical shell the net gravitational force is exactly zero, everywhere. A cavity at the center of a planet would be a weightless place, not because gravity switched off, but because the pulls from all directions cancel.

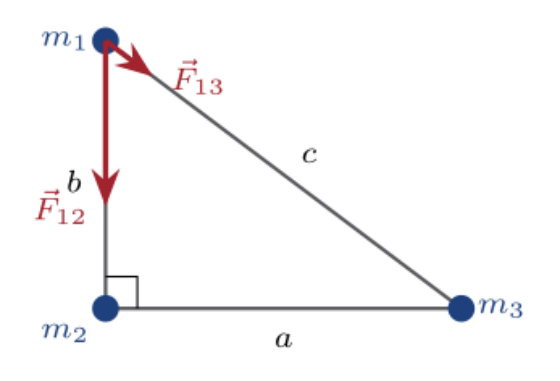
How was such a small constant ever measured? Not by astronomy, which only ever reveals the *product* Gm_1 for a body. In 1798 Henry Cavendish hung a light rod with two small lead spheres from a fine fiber, brought two large lead spheres alongside, and

measured the resulting twist. The forces involved were of order 10^{-7} N (comparable to the weight of a speck of dust), and the experiment had to be operated from an adjoining room by telescope so that the experimenter's own body heat would not stir the air. From that twist Cavendish extracted G, and with it the mass of the Earth. Two centuries of refinement have improved his value by only a few decimal places: G remains the least precisely known of the fundamental constants.

When several masses act on one body, the net gravitational force is the *vector* sum of the individual forces: the superposition principle. The next example shows the bookkeeping.

Example 13.1: Net gravitational force among three billiard balls

Three 0.170 kg billiard balls sit at the corners of a right triangle with legs $a = 0.400$ m and $b = 0.300$ m and hypotenuse $c = 0.500$ m. The ball m_2 is at the right-angle corner, with the cue ball m_1 a distance b from it and m_3 a distance a from it (so m_1 and m_3 are a distance c apart). Find the net gravitational force on m_1. Treat the balls as point masses at their centers, as the shell theorem permits.



•Figure description
Three masses m sub 1, m sub 2 and m sub 3 arranged with separations labeled a, b and c, so the net gravitational force on m sub 1 is the vector sum of the pulls from the other two.

Solution

Step 1: Coordinates and a common factor. Place m_2 at the origin, m_1 at $(0,\ 0.300)$, and m_3 at $(0.400,\ 0)$ (in meters). Each force is found from equation (13.1); with all masses equal to $m = 0.170$ kg, the product Gm_1m_j is the same for both pairs, so compute it once, squaring the mass first ($(0.170)^2 = 0.0289$):

$$Gm^2 = (6.674 \times 10^{-11}\,\mathrm{N\cdot m^2/kg^2})(0.170\,\mathrm{kg})^2 = 1.93 \times 10^{-12}\,\mathrm{N\cdot m^2}. \tag{13.1.2}$$

Step 2: The force from m_2. Gravity is attractive, so the force on m_1 points *toward* m_2, straight down. The distance is $b = 0.300$ m, and it is squared in the denominator:

$$F_{12} = \frac{Gm^2}{b^2} = \frac{1.93 \times 10^{-12}\,\mathrm{N\cdot m^2}}{(0.300\,\mathrm{m})^2} = \frac{1.93 \times 10^{-12}}{0.0900}\,\mathrm{N} = 2.14 \times 10^{-11}\,\mathrm{N} \quad (-y). \tag{13.1.3}$$

Step 3: The force from m_3. Its magnitude, at distance $c = 0.500$ m, is

$$F_{13} = \frac{Gm^2}{c^2} = \frac{1.93 \times 10^{-12}}{0.250}\,\mathrm{N} = 7.72 \times 10^{-12}\,\mathrm{N}. \tag{13.1.4}$$

It points from m_1 toward m_3. To get its direction, subtract the coordinates: from $(0,\ 0.300)$ to $(0.400,\ 0)$ is the displacement $(0.400,\ -0.300)$, of length 0.500 m. Dividing by that length gives the unit vector $(0.400/0.500,\ -0.300/0.500) = (0.8,\ -0.6)$. Multiply the magnitude by each component:

$$\mathbf{F}_{13} = F_{13}\,(0.8,\ -0.6) = (0.8 \times 7.72,\ -0.6 \times 7.72) \times 10^{-12}\,\mathrm{N} = (6.17,\ -4.63) \times 10^{-12}\,\mathrm{N}. \tag{13.1.5}$$

Step 4: Add components. The force from m_2 has no x-component, so

$$F_x = 6.17 \times 10^{-12}\,\mathrm{N},$$
$$F_y = -(21.4 + 4.63) \times 10^{-12}\,\mathrm{N} = -2.61 \times 10^{-11}\,\mathrm{N}. \tag{13.1.6}$$

(Writing 2.14×10^{-11} as 21.4×10^{-12} puts both terms over the same power of ten so they can be added.) The magnitude and direction are

$$F = \sqrt{F_x^2 + F_y^2} = \sqrt{(6.17)^2 + (26.1)^2} \times 10^{-12}\,\mathrm{N} \approx \boxed{2.7 \times 10^{-11}\,\mathrm{N}},$$
$$\theta = \arctan\frac{|F_y|}{F_x} = \arctan\frac{26.1}{6.17} \approx 77^\circ \text{ below the } +x\text{-axis}. \tag{13.1.7}$$

Notice how strongly the nearer ball dominates: m_2 is only $0.500/0.300 = 5/3$ times as close as m_3, but because distance enters squared it pulls $(5/3)^2 = 25/9 \approx 2.8$ times as hard. These forces are minuscule: a reminder of how weak gravity is between ordinary objects.

Checkpoint 13.1

The Sun is about 3×10^5 times more massive than the Earth, and yet you fall toward the Earth, not the Sun. Estimate the ratio of the Sun's pull on you to the Earth's pull on you, using $r_{\mathrm{Sun}} \approx 2.3 \times 10^4\ R_E$. Which force wins, and by roughly what factor? Then explain why you nevertheless do not feel the Sun's pull at all as you stand still. *(Hint for the last part: what else is being pulled by the Sun, and with what acceleration?)*

Answer

Answers to Checkpoints

Checkpoint 13.1. The ratio is $\dfrac{F_{\mathrm{Sun}}}{F_{\mathrm{Earth}}} = \dfrac{M_{\mathrm{Sun}}}{M_E}\left(\dfrac{R_E}{r_{\mathrm{Sun}}}\right)^2 \approx \dfrac{3.3 \times 10^5}{(2.3 \times 10^4)^2} = \dfrac{3.3 \times 10^5}{5.3 \times 10^8} \approx 6 \times 10^{-4}$. The Earth's pull wins by a factor of about 1600: the Sun's enormous mass is overwhelmed by the square of its enormous distance. You still do not feel even that small pull, because the Sun accelerates the Earth toward it at the same rate as it accelerates you. You and the ground beneath you fall toward the Sun together, so the Sun's pull produces no motion of you *relative* to the Earth and no force from the ground.

This page titled 13.1: Newton's Law of Universal Gravitation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.2: Applications of Universal Gravitation

Throughout this section and the next, one of the two bodies is enormously more massive than the other: a planet and a satellite, a star and a planet. We will consistently let m_1 denote the **large, central body** whose gravity sets the stage, and m_2 the small body moving in that gravity. As you are about to see, m_2 has a habit of canceling out.

Free-fall acceleration

Consider an object of mass m_2 a distance r from the center of a planet of mass m_1. In **free fall** (gravity being the only force acting) the gravitational attraction is the net force, so Newton's second law gives

$$m_2 a = G\,\frac{m_1 m_2}{r^2}. \tag{13.2.1}$$

The falling object's own mass appears on both sides and cancels. Writing g for this acceleration,

Key Equation: Free-fall acceleration

$$g = \frac{G m_1}{r^2}, \tag{13.2}$$

where r is the distance from the planet's center (so $r = R$ at the surface, and larger at altitude). The result depends on the planet's mass and on your distance from its center, but not at all on the mass of the object that is falling, which is Galileo's observation recovered as a theorem. This single formula gives the gravitational acceleration anywhere: on any planet, moon, or orbit.

As a first check, apply it to the Earth itself, with $m_E = 5.97 \times 10^{24}\ \text{kg}$ and $R_E = 6.37 \times 10^{6}\ \text{m}$: equation (13.2) gives $g = (6.674 \times 10^{-11})(5.97 \times 10^{24})/(6.37 \times 10^{6})^2 = 9.82\ \text{m/s}^2$, agreeing with the measured $9.81\ \text{m/s}^2$ to about a part in a thousand. The small residual comes mostly from the Earth's rotation and its slightly non-spherical shape. The product $Gm_E = 3.986 \times 10^{14}\ \text{m}^3/\text{s}^2$ will recur often enough to be worth remembering; astronomers know it far more precisely than they know G or m_E separately.

Example 13.2: Gravity on the Moon

The Moon has mass $m_M = 7.35 \times 10^{22}\ \text{kg}$ and radius $R_M = 1.74 \times 10^{6}\ \text{m}$. Find the free-fall acceleration at its surface.

Solution

Equation (13.2) applies with the Moon's mass and radius. Handle the numbers and the powers of ten separately. In the numerator, $6.674 \times 7.35 = 49.05$ and $10^{-11} \times 10^{22} = 10^{11}$, so $Gm_M = 49.05 \times 10^{11} = 4.91 \times 10^{12}$. In the denominator, squaring 1.74×10^{6} squares both parts: $(1.74)^2 = 3.03$ and $(10^6)^2 = 10^{12}$. Then

$$g = \frac{Gm_M}{R_M^2} = \frac{(6.674 \times 10^{-11})(7.35 \times 10^{22})}{(1.74 \times 10^{6})^2}\ \text{m/s}^2 = \frac{4.91 \times 10^{12}}{3.03 \times 10^{12}}\ \text{m/s}^2 \approx \boxed{1.62\ \text{m/s}^2}. \tag{13.2.2}$$

The powers of ten cancel exactly, leaving $4.91/3.03$. About one-sixth of Earth's $9.8\ \text{m/s}^2$ ($9.8/1.62 \approx 6.0$), which is why the Apollo astronauts could bound across the surface.

Example 13.3: Gravity at the Space Station

The International Space Station orbits at an altitude of about $400\ \text{km}$. Find the free-fall acceleration there. (Take $m_E = 5.97 \times 10^{24}\ \text{kg}$, $R_E = 6.37 \times 10^{6}\ \text{m}$.)

Solution

The r in equation (13.2) is the distance from Earth's *center*, not the altitude, so add the altitude to Earth's radius. Writing both in units of 10^6 m makes the addition easy, since $400\ \text{km} = 0.40 \times 10^6\ \text{m}$:

$$r = R_E + h = 6.37 \times 10^6 + 0.40 \times 10^6 = 6.77 \times 10^6\,\text{m}. \tag{13.2.3}$$

Then, with $Gm_E = (6.674 \times 10^{-11})(5.97 \times 10^{24}) = 3.986 \times 10^{14}\ \text{m}^3/\text{s}^2$,

$$g = \frac{Gm_E}{r^2} = \frac{3.986 \times 10^{14}}{(6.77 \times 10^6)^2}\,\text{m/s}^2 = \frac{3.986 \times 10^{14}}{4.58 \times 10^{13}}\,\text{m/s}^2 \approx \boxed{8.7\,\text{m/s}^2}. \tag{13.2.4}$$

Check. Compare with the surface by taking a ratio: $g_{\text{ISS}}/g_{\text{surf}} = (R_E/r)^2 = (6.37/6.77)^2 = (0.941)^2 = 0.885$, so $g_{\text{ISS}} = (0.885)(9.8) = 8.7\ \text{m/s}^2$. ✓ Gravity at the station is nearly 90% of its surface value: astronauts float not because gravity is absent but because they and the station are falling together.

Note: "Weightless" does not mean "gravity-free"

The previous example should be unsettling, because astronauts on the Space Station are visibly floating in a place where g is $8.7\ \text{m/s}^2$, almost as strong as on the ground. Gravity has not been escaped or canceled. What has vanished is the *contact* force: nothing is pushing up on the astronaut, because the floor is falling at exactly the same rate she is.

What a bathroom scale reads is never your weight $m_2 g$; it is the normal force the scale exerts on you, sometimes called your *apparent* weight. Stand still and the two agree. Ride a freely falling elevator, or orbit the Earth, and the normal force drops to zero while gravity carries on unchanged. Astronauts are not in "zero gravity"; they are in continual free fall, moving sideways fast enough that they keep missing the Earth. The correct term is **microgravity**, and the small residual it names comes from air drag and tidal effects, not from any weakening of g.

Checkpoint 13.2

A student reasons: "The Space Station is $400\ \text{km}$ up, and $400\ \text{km}$ is small compared with the Earth's $6370\ \text{km}$ radius, so g up there should be close to $9.8\ \text{m/s}^2$. But astronauts float, so g must actually be nearly zero. Both cannot be true." Identify precisely which step in the argument is wrong, and state the correct value of g at that altitude. Then answer this: at what distance from the Earth's center *would* g fall to one-hundredth of its surface value, and is any crewed spacecraft ever remotely that far out?

Answer

Circular orbits

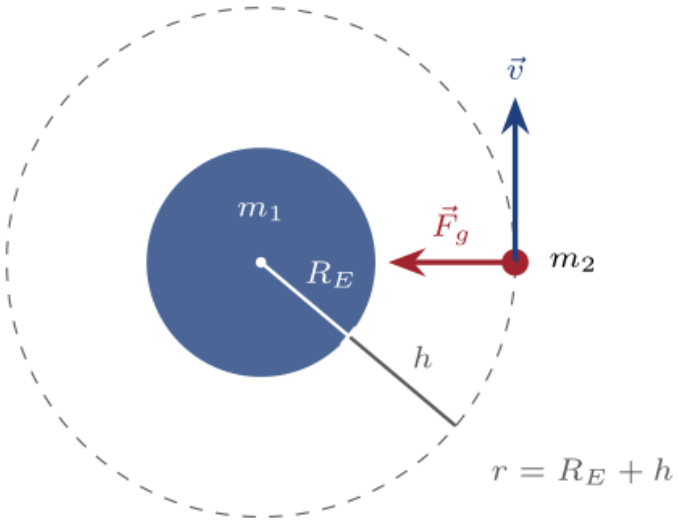


Figure 13.2. A satellite in a circular orbit. Gravity provides the centripetal force, $Gm_1m_2/r^2 = m_2v^2/r$, giving an orbital speed $v = \sqrt{Gm_1/r}$ that depends only on the orbital radius. The radius is measured from the planet's *center*, so for a satellite at altitude h above Earth, $r = R_E + h$.

•Figure description

A satellite of mass m sub 2 in a circular orbit around a central body of mass m sub 1. Its velocity is tangent to the orbit while the gravitational force points inward, supplying the centripetal force. A radial line from the planet's center to the orbit is split at the planet's surface: the inner part is the planet's radius R sub E and the outer part is the altitude h, so the orbital radius is r equals R sub E plus h.

A satellite in a circular orbit is held by gravity, which supplies exactly the centripetal force of Chapter 6. Setting the gravitational force equal to m_2v^2/r (Figure 13.2),

$$\frac{Gm_1m_2}{r^2} = \frac{m_2v^2}{r} \implies v = \sqrt{\frac{Gm_1}{r}}. \tag{13.2.5}$$

The satellite's mass cancels once again: the orbital speed depends only on the central mass and the orbital radius. A lower orbit is a faster one.

Example 13.4: Newton's Moon test

The chapter opened with Newton's claim that the Moon is falling exactly as an apple does, only sideways fast enough to keep missing. Test it. The Moon orbits at $r = 3.84 \times 10^8$ m with a period of $T = 27.3$ days. Compare the acceleration Newton's law predicts there with the acceleration the Moon's observed motion actually requires.

Solution

Predicted. Equation (13.2) says $g = Gm_E/r^2$. Write it once for the surface ($r = R_E$) and once for the Moon's orbit, and divide the second by the first; Gm_E cancels, leaving $g_{\text{Moon}}/g_{\text{surf}} = R_E^2/r^2 = (R_E/r)^2$. Multiply both sides by g_{surf}:

$$g_{\text{Moon}} = g_{\text{surf}}\left(\frac{R_E}{r}\right)^2 = (9.81)\left(\frac{6.37 \times 10^6}{3.84 \times 10^8}\right)^2 = (9.81)(0.0166)^2$$
$$[2pt] = (9.81)(2.75 \times 10^{-4}) = 2.70 \times 10^{-3}\,\text{m/s}^2. \tag{13.2.6}$$

The Moon is about 60 Earth radii away ($1/0.0166 \approx 60$), so the pull there is weaker by a factor of $60^2 = 3600$.

Observed. The Moon travels one circumference in one period. Convert the period to seconds first, $T = (27.3\text{ days})(86\,400\text{ s/day}) = 2.36 \times 10^6$ s. Then the orbital speed and the centripetal acceleration it requires are

$$v = \frac{2\pi r}{T} = \frac{2\pi(3.84\times10^{8}\,\mathrm{m})}{2.36\times10^{6}\,\mathrm{s}} = 1.02\times10^{3}\,\mathrm{m/s},$$
$$[2pt]a = \frac{v^2}{r} = \frac{(1.02\times10^{3}\,\mathrm{m/s})^2}{3.84\times10^{8}\,\mathrm{m}} = \frac{1.04\times10^{6}}{3.84\times10^{8}}\,\mathrm{m/s^2} = 2.72\times10^{-3}\,\mathrm{m/s^2}. \tag{13.2.7}$$

The two agree to about 1%, using nothing but the radius of the Earth, the distance to the Moon, and the length of the month. The same g that drops an apple, thinned by the inverse-square law across $384000\ \mathrm{km}$, is precisely what bends the Moon's path into a circle. This single comparison is what made gravity *universal*.

Example 13.5: A geosynchronous satellite

(a) Find the speed of a satellite in a circular orbit at altitude h above Earth, in terms of G, m_E, R_E, and h. (b) A **geosynchronous** satellite orbits with a period of one day. How fast is it moving, and how far from Earth's center? (Use $Gm_E = 3.986\times10^{14}\ \mathrm{m^3/s^2}$.)

Solution

(a) In a circular orbit, gravity supplies the centripetal force. Set the two equal, with $r = R_E + h$ the distance from Earth's center and m the satellite's mass:

$$\frac{Gm_E m}{r^2} = \frac{mv^2}{r}. \tag{13.2.8}$$

The satellite's mass cancels. Multiply both sides by r and take the square root:

$$v^2 = \frac{Gm_E}{r} \implies v = \sqrt{\frac{Gm_E}{r}} = \sqrt{\frac{Gm_E}{R_E + h}}. \tag{13.2.9}$$

Higher orbits are *slower*: r sits in the denominator.

(b) **Find r from the period.** One orbit is a distance $2\pi r$ covered at speed v, so $T = 2\pi r/v$. Substitute v from part (a). Dividing by a square root is multiplying by the square root of the reciprocal:

$$T = \frac{2\pi r}{\sqrt{Gm_E/r}} = 2\pi r\sqrt{\frac{r}{Gm_E}} = \frac{2\pi r^{3/2}}{\sqrt{Gm_E}}, \tag{13.2.10}$$

since $r\cdot\sqrt{r} = r^1\cdot r^{1/2} = r^{3/2}$. To solve for r, square both sides to clear the square roots, then multiply by Gm_E and divide by $4\pi^2$:

$$T^2 = \frac{4\pi^2 r^3}{Gm_E} \implies r^3 = \frac{Gm_E T^2}{4\pi^2}. \tag{13.2.11}$$

(This is Kepler's third law, which we meet formally in §13.4.) Finally, take the cube root (the $1/3$ power) of both sides. A day is $T = 86400\ \mathrm{s}$:

$$r = \left(\frac{Gm_E T^2}{4\pi^2}\right)^{1/3} = \left(\frac{(3.986\times10^{14})(86400)^2}{4\pi^2}\right)^{1/3}\mathrm{m}$$
$$= \left(7.54\times10^{22}\right)^{1/3}\mathrm{m} \approx 4.22\times10^{7}\,\mathrm{m}, \tag{13.2.12}$$

about $42000\ \mathrm{km}$ from Earth's center: some $4.22\times10^{7}/6.37\times10^{6} = 6.6$ Earth radii out.

The speed. Now part (a) gives

$$v = \sqrt{\frac{Gm_E}{r}} = \sqrt{\frac{3.986\times10^{14}\,\mathrm{m^3/s^2}}{4.22\times10^{7}\,\mathrm{m}}} = \sqrt{9.45\times10^{6}\,\mathrm{m^2/s^2}} \approx \boxed{3.07\times10^{3}\,\mathrm{m/s}}, \tag{13.2.13}$$

roughly $3.1\ \mathrm{km/s}$: the speed at which communications and weather satellites circle the Earth once per day.

Check. The speed should carry the satellite once around in a day: $2\pi r/v = 2\pi(4.22\times10^{7})/(3.07\times10^{3}) = 8.64\times10^{4}$ s, which is 86400 s. ✓

Note: Sidereal days, and what "geostationary" really means

Two refinements to the calculation above, both worth knowing. First, a satellite must keep pace with the stars rather than the Sun, so the period to match is the *sidereal* day, 86164 s rather than 86400 s; using it lowers r by about 80 km. Second, a *geosynchronous* orbit is any orbit with a one-day period, and such a satellite generally traces a slow figure-eight in the sky. Only if the orbit is also circular *and* directly above the equator does the satellite hang truly motionless overhead: that special case is called **geostationary**, and it is what a fixed satellite dish points at.

Checkpoint 13.3

A cargo capsule is in a circular orbit and needs to catch up with a space station ahead of it in the *same* orbit. The pilot fires the thrusters forward, along the direction of motion, to speed up. Use $v = \sqrt{Gm_1/r}$ to explain why this maneuver causes the capsule to end up *behind* where it would otherwise have been, and what the pilot should have done instead. *(Hint: adding energy raises the orbit; what does a larger r do to the period?)*

Answer

Answers to Checkpoints

Checkpoint 13.2. The wrong step is "astronauts float, so g must be nearly zero." Gravity at the station is strong, $g = 9.8\,(6.37/6.77)^2 \approx 8.7$ m/s^2. The astronauts float because they and the station are *falling together* around the Earth, so nothing presses on them; that is weightlessness in free fall, not absence of gravity. Since $g \propto 1/r^2$, g falls to one-hundredth of its surface value at $r = 10R_E \approx 64000$ km from the center. Crewed spacecraft in Earth orbit never come close to that. The only people ever to travel that far are the Apollo astronauts and, in 2026, the Artemis II crew, all on their way to the Moon, about $60R_E$ away.

Checkpoint 13.3. Firing forward adds energy, and a capsule with more energy climbs into a higher orbit. But in a higher orbit $v = \sqrt{Gm_1/r}$ is *smaller*, and the period, which grows as $r^{3/2}$, is longer, so the capsule falls behind the station. To catch up, the pilot should fire *backward*: that drops the capsule into a lower, faster orbit with a shorter period, in which it gains on the station, and then fire forward at the right moment to rise back to the station's orbit.

This page titled 13.2: Applications of Universal Gravitation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.3: Gravitational Potential Energy

In Chapter 8 we wrote gravitational potential energy as $U = m_2 g h$, but that assumed a *constant* g, valid only near a planet's surface. Over astronomical distances g falls off as $1/r^2$, and the potential energy must be found by integrating the true gravitational force. The result, taking $U = 0$ at infinite separation, is

Key Equation: Gravitational potential energy

$$U = -\frac{G m_1 m_2}{r}. \tag{13.3}$$

The potential energy is *negative*, approaching zero as $r \to \infty$ and growing more negative as the masses draw together: reflecting that gravity is attractive and that energy must be *added* to pull two masses apart. Note that U belongs to the *pair* of bodies, not to either one alone. As long as gravity is the only force, the total mechanical energy $E = K + U$ is conserved, exactly as in Chapter 8.

Note: Where did $U = m_2 g h$ go?

Equation (13.3) looks nothing like $U = m_2 g h$: one is negative and shrinks with distance, the other is positive and grows with height. They are the same statement, because only *changes* in potential energy have physical meaning. Lift a mass from the surface R to a height h above it:

$$\Delta U = -\frac{G m_1 m_2}{R + h} + \frac{G m_1 m_2}{R} = G m_1 m_2 \, \frac{h}{R(R + h)}. \tag{13.3.1}$$

For $h \ll R$ the denominator is essentially R^2, so

$$\Delta U \approx \frac{G m_1}{R^2} \, m_2 h = m_2 g h. \tag{13.3.2}$$

The familiar formula is the near-surface limit of the exact one, and the choice $U = 0$ at infinity is simply a different (and, for orbital problems, far more convenient) zero point than $U = 0$ at the floor. Use $m_2 g h$ for ladders and roller coasters; use $-G m_1 m_2 / r$ the moment r changes by an appreciable fraction of R.

For a satellite in a circular orbit, combining $K = \tfrac{1}{2} m_2 v^2$ with $v^2 = G m_1 / r$ gives the tidy total energy

$$E = K + U = \frac{G m_1 m_2}{2r} - \frac{G m_1 m_2}{r} = -\frac{G m_1 m_2}{2r}. \tag{13.4}$$

The energy is negative: the hallmark of a *bound* orbit. Note also that $K = -E$ and $U = 2E$: a circular orbit's kinetic energy is exactly half the magnitude of its potential energy, a relation that turns up throughout astrophysics.

To break free entirely, a body must be given enough energy to reach $E = 0$ (just barely escaping, arriving at infinity with zero speed). Setting $\tfrac{1}{2} m_2 v_{\text{esc}}^2 - G m_1 m_2 / R = 0$, with R the radius at which it is launched, yields the **escape speed**

Key Equation: Escape speed

$$v_{\text{esc}} = \sqrt{\frac{2 G m_1}{R}}, \tag{13.5}$$

which for Earth is about 11.2 km/s. Two features are worth pausing over. It is independent of the escaping object's mass: a pebble and a spacecraft need the same speed. And because energy is a scalar, it is independent of *direction* as well: fired straight up, sideways, or at 45°, anything leaving Earth's surface at 11.2 km/s escapes, provided it does not run into the ground on the way. Comparing equations (13.5) and $v = \sqrt{G m_1 / r}$ gives a relation worth memorizing: at any radius,

$$v_{\text{esc}} = \sqrt{2} \, v_{\text{orbit}}. \tag{13.3.3}$$

Escaping costs only about 41% more speed than circling.

Example 13.6: Speed of an incoming asteroid

An asteroid heading directly toward Earth has a speed of $12.0\ \text{km/s}$ when it is 100 Earth radii from Earth's center. How fast is it moving (a) when it passes the Moon's distance, about 60 Earth radii from center, and (b) when it reaches the Earth's surface? (Take $Gm_E = 3.986 \times 10^{14}\ \text{m}^3/\text{s}^2$, $R_E = 6.37 \times 10^6\ \text{m}$; ignore the atmosphere.)

Solution

Step 1: Energy conservation. Only gravity acts, so mechanical energy is conserved. With the gravitational potential energy $U = -Gm_E m/r$, and the asteroid's mass m canceling from every term,

$$\tfrac{1}{2}v_1^2 - \frac{Gm_E}{r_1} = \tfrac{1}{2}v_2^2 - \frac{Gm_E}{r_2}. \tag{13.3.4}$$

To solve for v_2^2, add Gm_E/r_2 to both sides, then multiply every term by 2:

$$v_2^2 = v_1^2 + \frac{2Gm_E}{r_2} - \frac{2Gm_E}{r_1} = v_1^2 + 2Gm_E\left(\frac{1}{r_2} - \frac{1}{r_1}\right). \tag{13.3.5}$$

Because the asteroid moves inward, $r_2 < r_1$, so $1/r_2 > 1/r_1$ and the bracket is positive: the asteroid speeds up, as it must.

Step 2: Part (a). With $r_1 = 100R_E = 6.37 \times 10^8\ \text{m}$ and $r_2 = 60R_E = 3.82 \times 10^8\ \text{m}$, evaluate the bracket term on its own first:

$$2Gm_E\left(\frac{1}{r_2} - \frac{1}{r_1}\right) = 2(3.986 \times 10^{14})\left(\frac{1}{3.82 \times 10^8} - \frac{1}{6.37 \times 10^8}\right) = 8.34 \times 10^5\,\text{m}^2/\text{s}^2. \tag{13.3.6}$$

The initial speed must be in m/s, $v_1 = 12.0 \times 10^3\ \text{m/s}$, so $v_1^2 = 1.440 \times 10^8\ \text{m}^2/\text{s}^2$. Then

$$v_2^2 = 1.440 \times 10^8 + 0.00834 \times 10^8 = 1.448 \times 10^8\,\text{m}^2/\text{s}^2 \implies v_2 \approx \boxed{12.03\,\text{km/s}}. \tag{13.3.7}$$

Step 3: Part (b). Now take $r_2 = R_E = 6.37 \times 10^6\ \text{m}$. The bracket contributes $1.239 \times 10^8\ \text{m}^2/\text{s}^2$, almost as much as the initial v_1^2, and

$$v_2^2 = 1.44 \times 10^8 + 1.239 \times 10^8 = 2.679 \times 10^8\,\text{m}^2/\text{s}^2 \implies v_2 \approx \boxed{16.4\,\text{km/s}}. \tag{13.3.8}$$

Between $100R_E$ and $60R_E$ the asteroid speeds up by a mere 0.3%: out there Earth's gravity is feeble, and the kinetic energy it adds is nothing next to what the asteroid already has. Almost all of the speeding-up happens in the last stretch, because $1/r$ grows fastest at small r.

Check. An object falling in from very far away arrives with $v = \sqrt{v_\infty^2 + v_{\text{esc}}^2}$ (the energy equation with $1/r_1 \to 0$, since $2Gm_E/R_E = v_{\text{esc}}^2$). With $v_{\text{esc}} = 11.2\ \text{km/s}$, $\sqrt{(12.0)^2 + (11.2)^2} = \sqrt{144 + 125} = 16.4\ \text{km/s}$. ✓ Starting at $100R_E$ rather than infinity makes almost no difference. Nothing can strike the Earth at less than escape speed.

Checkpoint 13.4

Rockets are launched eastward, and from sites as near the equator as possible. They are also launched to the *east* rather than straight up, tipping over within a minute or two of liftoff. Using equation (13.5) and the fact that escape speed does not depend on direction, explain what each of these choices actually buys, and explain why, despite direction being irrelevant to escape speed, "straight up" is nevertheless a terrible way to reach orbit.

Answer

Answers to Checkpoints

Checkpoint 13.4. Escape speed is set by energy, $v_{\text{esc}} = \sqrt{2Gm_E/R_E}$, and energy does not care about direction; but a rocket's speed *relative to space* does include the Earth's own rotation. Launching eastward adds the ground's rotational speed (about 465

$\mathrm{m/s}$ at the equator) for free, and launching near the equator makes that bonus as large as possible. “Straight up” is terrible for reaching orbit because orbit is not about height; it requires a large *horizontal* speed, about $7.8\ \mathrm{km/s}$ in low orbit. A rocket that climbed straight up would gain altitude and then fall back, and all the while it would be spending thrust simply holding itself up against gravity. Tipping over lets the engines spend their thrust on the horizontal speed an orbit needs.

This page titled 13.3: Gravitational Potential Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.4: Kepler's Laws of Planetary Motion

Decades before Newton, Johannes Kepler distilled Tycho Brahe's lifetime of planetary observations into three empirical laws. Newton's gravity explains all three. Although Kepler stated them for planets orbiting the Sun, they hold for any body orbiting a much more massive one: moons about planets, satellites about the Earth, stars about the center of the galaxy.

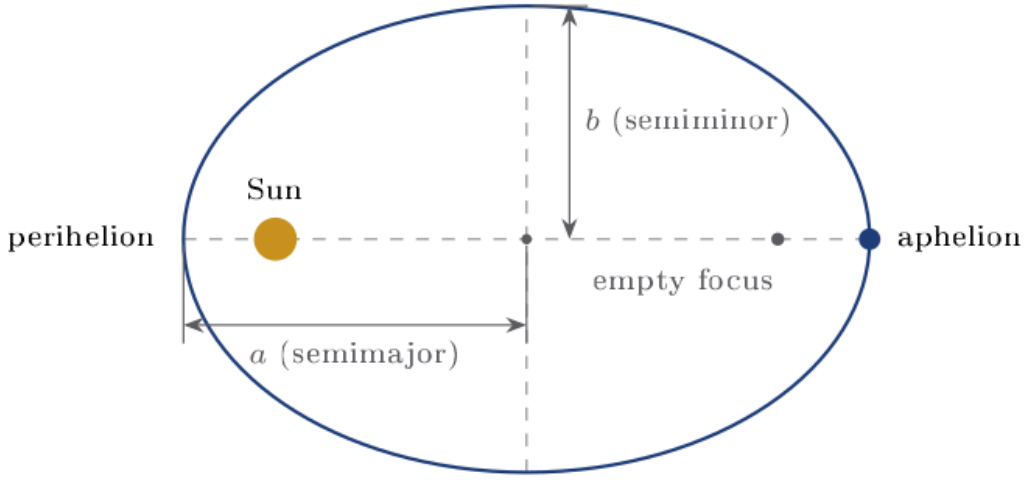


Figure 13.3. Kepler's first law: a planet's orbit is an ellipse with the Sun at one focus. The semimajor axis a is half the long axis and the semiminor axis b is half the short axis; perihelion and aphelion are the nearest and farthest points. The eccentricity is exaggerated here: Earth's orbit is so nearly circular that on this scale the two foci would be almost indistinguishable.

•Figure description

An elliptical planetary orbit with the Sun at one focus and the other focus empty. The semimajor axis a is marked as half the long axis and the semiminor axis b as half the short axis, with perihelion the closest point to the Sun and aphelion the farthest.

Kepler's first law: *Every planet moves in an elliptical orbit with the Sun at one focus* (Figure 13.3). A circle is the special case in which the two foci coincide; real orbits are slightly elongated, with the planet nearest the Sun at **perihelion** and farthest at **aphelion**. Nothing physical sits at the second focus: it is a geometric consequence of the inverse-square law, nothing more.

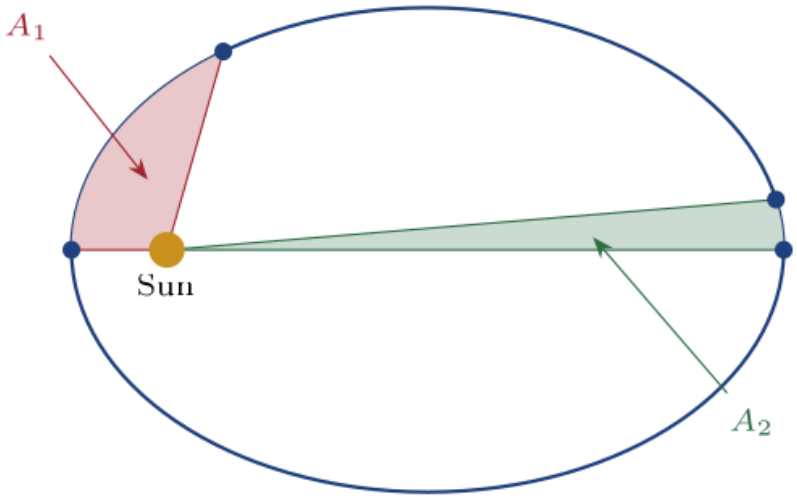


Figure 13.4. Kepler's second law: the Sun–planet line sweeps equal areas in equal times, $A_1 = A_2$. Near the Sun the planet swings through a wide angle in that time; far away it creeps through a narrow one. The planet therefore moves fastest at perihelion and slowest at aphelion.

•Figure description

An elliptical orbit with two shaded wedges swept out by the Sun-planet line, labeled A sub 1 near the Sun and A sub 2 far from it. The two areas are equal, swept in equal times, so the planet moves faster when nearer the Sun.

Kepler's second law: *The line from the Sun to a planet sweeps out equal areas in equal times* (Figure 13.4). This means a planet moves *faster* when near the Sun and *slower* when far away. It is a direct consequence of the conservation of angular momentum from Chapter 11: gravity pulls straight along the Sun–planet line and so exerts no torque about the Sun, leaving L constant. Quantitatively, the rate at which area is swept is

$$\frac{dA}{dt} = \frac{L}{2m_2}, \tag{13.4.1}$$

and since L never changes, neither does dA/dt, which is Kepler's second law exactly. Combined with $L = m_2 v r$ at perihelion and aphelion (where $\mathbf{v} \perp \mathbf{r}$), this gives the useful relation $v_p r_p = v_a r_a$.

Kepler's third law: *The square of a planet's orbital period is proportional to the cube of its semimajor axis*. For a circular orbit this follows directly from $v = \sqrt{Gm_1/r}$ and $T = 2\pi r/v$, which give:

Key Equation: Kepler's third law

$$T^2 = \frac{4\pi^2}{Gm_1}\, a^3, \tag{13.6}$$

where a is the semimajor axis (the orbital radius for a circle) and m_1 is the central mass. Although we derived it for a circle, the law holds for every ellipse once r is replaced by a. The proportionality constant depends only on m_1, so a single measured orbit reveals the mass it circles, and it does so without any knowledge of the orbiting body's mass, which is why we can weigh a star we have never visited by watching a planet go around it.

Two footnotes on that. Strictly, the constant involves the *total* mass $m_1 + m_2$, because both bodies orbit their common center of mass; for a planet around a star, or a satellite around a planet, m_2 is negligible and equation (13.6) is excellent. And the orbital energy result of equation (13.4) generalizes the same way: an elliptical orbit has $E = -Gm_1 m_2/2a$, depending on the semimajor axis alone and not on how elongated the orbit is.

Example 13.7: Weighing the Sun

The Earth orbits the Sun once every $T = 3.156 \times 10^7$ s at a mean distance $a = 1.496 \times 10^{11}$ m. Find the mass of the Sun.

Solution

Kepler's third law for a planet of orbit radius a around a central mass m_S reads $T^2 = 4\pi^2 a^3/(Gm_S)$. We want m_S, which sits in the denominator: multiply both sides by Gm_S, then divide both sides by GT^2:

$$Gm_S T^2 = 4\pi^2 a^3 \implies m_S = \frac{4\pi^2 a^3}{GT^2}. \tag{13.4.2}$$

Evaluate numerator and denominator separately. Cubing 1.496×10^{11} cubes both parts: $(1.496)^3 = 3.348$ and $(10^{11})^3 = 10^{33}$, so $4\pi^2 a^3 = (39.48)(3.348 \times 10^{33}) = 1.32 \times 10^{35}$. In the denominator, $(3.156 \times 10^7)^2 = 9.960 \times 10^{14}$, and $(6.674 \times 10^{-11})(9.960 \times 10^{14}) = 6.65 \times 10^4$. So

$$m_S = \frac{4\pi^2 (1.496 \times 10^{11})^3}{(6.674 \times 10^{-11})(3.156 \times 10^7)^2} = \frac{1.32 \times 10^{35}}{6.65 \times 10^4}\ \text{kg} \approx \boxed{1.99 \times 10^{30}\ \text{kg}}. \tag{13.4.3}$$

(Dividing powers of ten subtracts exponents: $10^{35}/10^4 = 10^{31}$, and $1.32/6.65 = 0.199$, giving $0.199 \times 10^{31} = 1.99 \times 10^{30}$.) The units reduce to kilograms: $\text{m}^3/(\text{N}\cdot\text{m}^2/\text{kg}^2 \cdot \text{s}^2) = \text{m}^3 \cdot \text{kg}^2/(\text{kg}\cdot\text{m}^3) = \text{kg}$.

We have weighed the Sun using nothing but the size and period of Earth's orbit: and, notably, without ever needing the Earth's own mass. The Sun turns out to be about 3×10^5 times as massive as the Earth.

Checkpoint 13.5

Jupiter is about 5.2 times as far from the Sun as the Earth is. Without a calculator, estimate Jupiter's orbital period in Earth years, and check your estimate against the fact that Jupiter takes roughly 12 years to circle the sky once. Then answer a question Kepler could not: two of Jupiter's moons have very different masses but happen to orbit at the same radius. Which one has the longer period?

Answer

Answers to Checkpoints

Checkpoint 13.5. By Kepler's third law, $T^2 \propto a^3$, so $T_J = (5.2)^{3/2}$ years. Since $5.2^{3/2} = 5.2 \times \sqrt{5.2} \approx 5.2 \times 2.3 \approx 12$, Jupiter's period is about 12 years, matching the observation. The two moons have the *same* period: $T^2 = 4\pi^2 a^3/(GM)$ depends only on the orbit's size and the central mass (Jupiter's), not on the orbiting body's own mass.

This page titled 13.4: Kepler's Laws of Planetary Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.S: Universal Gravitation (Summary)

Summary

- **Newton's law of universal gravitation**: every pair of masses attracts with $F_g = Gm_1m_2/r^2$, directed along the line joining them ($G = 6.674 \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$), with r measured center to center. Several gravitational forces add as vectors.
- By the **shell theorem**, a spherically symmetric body pulls on anything outside it exactly as a point mass at its center would, which is what licenses using center-to-center distances for planets and stars.
- The **free-fall acceleration** at distance r from a body of mass m_1 is $g = Gm_1/r^2$. A satellite in a circular orbit has speed $v = \sqrt{Gm_1/r}$. Both are independent of the falling or orbiting mass.
- **Gravitational potential energy** is $U = -Gm_1m_2/r$ (zero at infinity, negative for bound systems), reducing to m_2gh near a surface. A circular orbit has total energy $E = -Gm_1m_2/(2r)$, and the escape speed is $v_{\mathrm{esc}} = \sqrt{2Gm_1/R} = \sqrt{2}\ v_{\mathrm{orbit}}$.
- **Kepler's laws**: orbits are ellipses with the Sun at a focus (first); the Sun–planet line sweeps equal areas in equal times, from angular-momentum conservation (second); and $T^2 = (4\pi^2/Gm_1)\ a^3$ (third), which lets a single orbit reveal the central mass.

Key Equations

Newton's law of universal gravitation

$$F_g = G\,\frac{m_1 m_2}{r^2} \tag{13.S.1}$$

Free-fall acceleration

$$g = \frac{Gm_1}{r^2} \tag{13.S.2}$$

Gravitational potential energy

$$U = -\frac{Gm_1m_2}{r} \tag{13.S.3}$$

Escape speed

$$v_{\mathrm{esc}} = \sqrt{\frac{2Gm_1}{R}} \tag{13.S.4}$$

Kepler's third law

$$T^2 = \frac{4\pi^2}{Gm_1}\,a^3 \tag{13.S.5}$$

This page titled 13.S: Universal Gravitation (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

13.E: Universal Gravitation (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 13.1 — Newton's Law of Universal Gravitation

13.1 ★ Two $75\ \text{kg}$ students sit $1.0\ \text{m}$ apart, center to center. Treating them as point masses, find the gravitational force between them, and compare it with the weight of a single grain of sand (about $1 \times 10^{-5}\ \text{N}$).

13.2 ★ *[Conceptual.]* The gravitational force between two objects follows an inverse-square law.

a. If the distance between two masses is tripled, by what factor does the force change?
b. Explain the geometric reason an inverse-square law arises for an influence spreading from a point.

13.3 ★★ Two lead spheres, $6.0\ \text{kg}$ and $12\ \text{kg}$, have centers $0.25\ \text{m}$ apart. Find the gravitational force between them.

13.4 ★★★ Two $2.0\ \text{kg}$ masses sit at $(0, 0)$ and $(0.60, 0)$ m. A third $2.0\ \text{kg}$ mass sits at $(0.30, 0.40)$ m. Find the magnitude and direction of the net gravitational force on the third mass.

13.5 ★ *[Conceptual.]* When an apple falls from a tree, the Earth pulls the apple down.

a. Does the apple pull the Earth up with a force of the same size? Justify using Newton's third law and equation (13.1).
b. If the forces are equal, why does the apple accelerate noticeably while the Earth does not?

Section 13.2 — Applications of Universal Gravitation

13.6 ★ Mars has mass $6.42 \times 10^{23}\ \text{kg}$ and radius $3.39 \times 10^{6}\ \text{m}$. Find the free-fall acceleration at its surface.

13.7 ★★ *[Real-world: spaceflight.]* A satellite orbits the Earth in a circular orbit at an altitude of $500\ \text{km}$. (Take $m_E = 5.97 \times 10^{24}\ \text{kg}$, $R_E = 6.37 \times 10^{6}\ \text{m}$.)

a. Find its orbital speed.
b. Find its orbital period.

13.8 ★★ At what altitude above Earth's surface is the free-fall acceleration half its surface value? Express your answer as a multiple of R_E.

13.9 ★★ *[Conceptual.]* Astronauts aboard the orbiting Space Station appear weightless, yet the free-fall acceleration there is nearly $8.7\ \text{m/s}^2$. Explain why they float, and why "zero gravity" is a misleading description.

13.10 ★★★ A planet of mass m_1 has a moon in a circular orbit of radius r and period T. Show that the planet's mass can be written $m_1 = 4\pi^2 r^3/(GT^2)$, and explain how this lets astronomers weigh distant planets.

Section 13.3 — Gravitational Potential Energy

13.11 ★ Find the gravitational potential energy of a $1500\ \text{kg}$ spacecraft and the Earth when the craft is at an altitude equal to one Earth radius (so $r = 2R_E$). (Take $m_E = 5.97 \times 10^{24}\ \text{kg}$, $R_E = 6.37 \times 10^{6}\ \text{m}$.)

13.12 ★ *[Conceptual.]* Gravitational potential energy $U = -Gm_1m_2/r$ is always negative.

a. What is the physical meaning of the negative sign?
b. What does it mean for a system to have total energy $E < 0$ versus $E \geq 0$?

13.13 ★★ *[Real-world: spaceflight.]* Find the escape speed from the surface of the Moon ($m_M = 7.35 \times 10^{22}\ \text{kg}$, $R_M = 1.74$

$\times 10^6$ m), and compare it with Earth's 11.2 km/s.

13.14 ★★ A 2000 kg satellite is in a circular orbit at radius $r = 2R_E$.

a. Find its kinetic energy.
b. Find its gravitational potential energy.
c. Find its total mechanical energy, and comment on the sign.

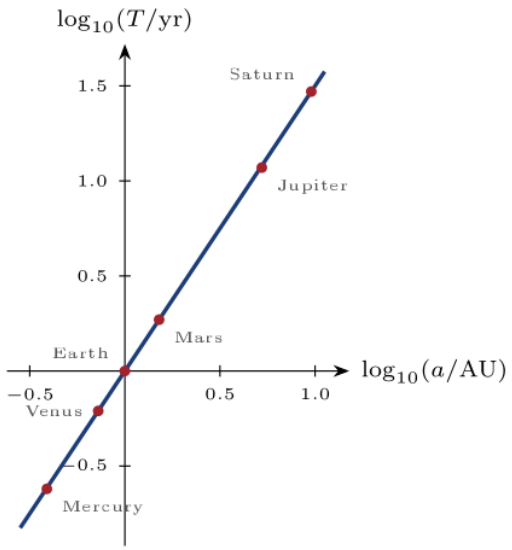
13.15 ★★★ A projectile is launched straight up from Earth's surface at 8.0 km/s (less than escape speed). Using energy conservation, find the maximum distance from Earth's center it reaches. (Take $m_E = 5.97 \times 10^{24}$ kg, $R_E = 6.37 \times 10^6$ m.)

Section 13.4 — Kepler's Laws of Planetary Motion

13.16 ★ *[Conceptual.]* Using Kepler's second law, explain why Earth moves fastest in early January (when it is closest to the Sun) and slowest in early July. Which conservation law underlies this?

13.17 ★★ Mars orbits the Sun at a mean distance of 1.52 AU (where 1 AU is Earth's orbital radius). Use Kepler's third law to find Mars's orbital period in Earth years, without needing the Sun's mass.

13.18 ★★ *[Graph-reading.]* The plot shows orbital period versus semimajor axis for six of the Sun's planets, on logarithmic axes, with distances in AU and periods in Earth years. Both axes use the same scale, so slopes may be read directly off the grid.



•Figure description
A logarithmic plot of orbital period in Earth years against semimajor axis in astronomical units for the planets, from Mercury and Venus outward. The points fall along a straight line whose slope of three halves expresses Kepler's third law.

a. Kepler's third law $T^2 \propto a^3$ predicts a straight line on a log–log plot. What slope should the line have?
b. Measure the slope from the graph using any two convenient points, and compare.
c. Explain why the line passes through the origin, and what the intercept would have been had the periods been measured in days instead of years.

13.19 ★★ A comet orbits the Sun with a period of 76 years (like Halley's comet). Find the semimajor axis of its orbit in AU. *(Use Kepler's third law with the Earth as a reference: $T^2 = a^3$ when T is in years and a in AU.)*

13.20 ★★★ *[Real-world: astronomy.]* One of Jupiter's moons, Io, orbits at 4.22×10^8 m with a period of 1.53×10^5 s (about 1.77 days). Use Kepler's third law to determine the mass of Jupiter.

13.21 ★★★ *[Real-world: astronomy.]* The Earth ($m_E = 5.97 \times 10^{24}$ kg) and Moon ($m_M = 7.35 \times 10^{22}$ kg) are separated, center to center, by 3.84×10^8 m. A spacecraft travels along the line joining them.

a. Write the condition for the net gravitational force on the spacecraft to be zero.
b. Find the distance of this null point from the center of the Earth.
c. Express it as a fraction of the Earth–Moon distance, and give the distance from the Moon.
d. Explain physically why the null point lies so much closer to the Moon than to the Earth.

13.22 ★★★ *[Real-world: spaceflight.]* A $500\ \text{kg}$ satellite is in a circular orbit at $r_1 = 6.67 \times 10^6\ \text{m}$ (low Earth orbit) and is to be raised to geostationary radius $r_2 = 4.22 \times 10^7\ \text{m}$. (Use $Gm_E = 3.986 \times 10^{14}\ \text{m}^3/\text{s}^2$.)

a. Find the orbital speed and the total mechanical energy in the low orbit.
b. Find the total mechanical energy in the geostationary orbit.
c. How much energy must be added to raise the satellite between these orbits?
d. The higher orbit has more total energy, yet the satellite moves *slower* there. Resolve this apparent paradox.

13.23 ★★★ *[Real-world: astronomy.]* A comet orbits the Sun ($m_S = 1.99 \times 10^{30}\ \text{kg}$). At perihelion it is $r_p = 8.9 \times 10^{10}\ \text{m}$ from the Sun; at aphelion, $r_a = 5.3 \times 10^{12}\ \text{m}$.

a. Use conservation of angular momentum to relate the perihelion and aphelion speeds.
b. Use energy conservation to find the comet's speed at perihelion, and then at aphelion.
c. Compare the perihelion speed with the speed of a circular orbit at that same distance. Explain, using $v_{\text{esc}} = \sqrt{2}\ v_{\text{orbit}}$, why your answer had to come out close to (but below) $\sqrt{2}$.

This page titled 13.E: Universal Gravitation (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14: FLUIDS

Until now our objects have been rigid: solid blocks, rods, and spheres that hold their shape. But much of the world is made of **fluids**: substances that flow and take the shape of their container. Water, air, blood, oil, and even the slowly creeping rock of the Earth's mantle are fluids. As Bruce Lee put it, "Be like water": shapeless, formless, filling whatever holds it.

Fluids obey the same Newtonian mechanics as everything else, but because they deform and flow, we describe them with new quantities (**density** and **pressure**) rather than tracking individual particles. From these grow a remarkable set of principles: that pressure increases with depth, that a confined fluid transmits pressure undiminished (Pascal), that a submerged body is buoyed up by the weight of fluid it displaces (Archimedes), and that a moving fluid trades pressure for speed (Bernoulli). Together they explain why ships float, how hydraulic brakes multiply force, and how airplanes fly.

Learning Objectives

- Define density and pressure, and distinguish absolute pressure from gauge pressure.
- Calculate how pressure varies with depth, and apply Pascal's principle to hydraulic systems.
- Apply Archimedes' principle to buoyancy and flotation, and predict what floats and how deep.
- Use the continuity equation and Bernoulli's equation to analyze fluids in motion, and recognize where they stop applying.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

14.1: Fluids, Density, and Pressure
14.2: Pressure as a Function of Depth
14.3: Pascal's Principle
14.4: Buoyancy and Archimedes' Principle
14.5: Fluid Dynamics and Bernoulli's Principle
14.S: Fluids (Summary)
14.E: Fluids (Problems)

This page titled 14: Fluids is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.1: Fluids, Density, and Pressure

A **fluid** is any substance that flows and conforms to the shape of its container, both liquids and gases. To describe how much matter a fluid packs into a region, we use its **density**, the mass per unit volume:

Key Equation: Density

$$\rho = \frac{m}{V}. \tag{14.1}$$

Density is measured in $\mathrm{kg/m^3}$. Water has $\rho = 1000\ \mathrm{kg/m^3}$, while air at $0\,^\circ\mathrm{C}$ and $1\ \mathrm{atm}$ is about $1.29\ \mathrm{kg/m^3}$: some 800 times less. That qualifier matters: a liquid's density barely budges with pressure or temperature, but a gas compresses readily, so a gas density is meaningless unless the conditions are stated (the same air at $20\,^\circ\mathrm{C}$ has $\rho \approx 1.2\ \mathrm{kg/m^3}$). Differences in density are what make oil float on water and a hot-air balloon rise.

When a fluid pushes on a surface, what matters is not just the force but how it is spread out. **Pressure** is the force per unit area exerted perpendicular to a surface:

Key Equation: Pressure

$$P = \frac{F}{A}. \tag{14.2}$$

The SI unit is the **pascal** ($\mathrm{Pa} = \mathrm{N/m^2}$). A pascal is tiny (a dollar bill lying flat presses on a table with about $1\ \mathrm{Pa}$), so real pressures are usually quoted in kilopascals or in one of several older units. One chain of equivalents is worth committing to memory:

$$1\,\mathrm{atm} = 1.013 \times 10^5\,\mathrm{Pa} = 760\,\mathrm{mmHg} = 14.7\,\mathrm{psi}, \tag{14.1.1}$$

where psi is pounds per square inch (the unit on a tire gauge) and mmHg, the millimeter of mercury, is also called the torr; blood pressures are reported in it. The same force spread over a larger area gives less pressure, which is exactly why you can lie on a bed of nails unharmed: your weight is divided among so many points that the pressure at any one is bearable.

One feature of pressure trips students up. We defined it using the force on a surface, but pressure at a point inside a fluid has no direction of its own: a small surface placed at that point feels the same push whichever way it faces. *Pressure is a scalar.* What is a vector is the force it produces, which is always perpendicular to the surface and directed into it. This is why water pushes sideways on the face of a dam and upward on the underside of a boat's hull, not merely downward.

Example 14.1: Pressure under a heel

A $55\ \mathrm{kg}$ woman balances all her weight on the tip of one stiletto heel of area $1.0\ \mathrm{cm^2}$. Find the pressure on the floor, and compare it with a flat shoe of area $150\ \mathrm{cm^2}$.

Solution

Step 1: The force. The force on the floor is her weight, $F = mg = (55\ \mathrm{kg})(9.8\ \mathrm{m/s^2}) = 539\ \mathrm{N}$.

Step 2: Convert the area to SI. Pressure in pascals needs the area in square meters. Since $1\ \mathrm{cm} = 10^{-2}\ \mathrm{m}$, squaring both sides gives $1\ \mathrm{cm^2} = (10^{-2})^2\ \mathrm{m^2} = 10^{-4}\ \mathrm{m^2}$; the conversion factor must be squared along with the unit, exactly as in Chapter 1. So the heel's area is $A = 1.0 \times 10^{-4}\ \mathrm{m^2}$.

Step 3: The pressure.

$$P = \frac{F}{A} = \frac{539\,\mathrm{N}}{1.0 \times 10^{-4}\,\mathrm{m^2}} \approx \boxed{5.4 \times 10^6\,\mathrm{Pa}}. \tag{14.1.2}$$

Dividing by 10^{-4} is the same as multiplying by 10^{4}, so the pressure is $539 \times 10^{4} = 5.39 \times 10^{6}\ \text{Pa}$. That is $5.39 \times 10^{6} / 1.01 \times 10^{5} \approx 53$ times atmospheric pressure: enough to dent a wooden floor.

The flat shoe. Spread over $150\ \text{cm}^2 = 150 \times 10^{-4}\ \text{m}^2$, the same force gives $P = 539/(1.50 \times 10^{-2}) = 3.6 \times 10^{4}\ \text{Pa}$. Because $P = F/A$ with F fixed, pressure is inversely proportional to area: 150 times the area means $1/150$ the pressure. Same force, vastly different pressure.

Checkpoint 14.1

A woman in flat shoes and an African elephant stand side by side on soft ground. Decide which is more likely to sink in, and state exactly what you would need to look up to be sure. Then sharpen it: an elephant weighs roughly 60 times what the woman does, but spreads that weight over four feet whose total contact area is only about 20 times larger than her two soles. By what factor do the two pressures differ, and which way? *(Hint: you are comparing F/A, so the two ratios do not cancel; they compete.)*

Answer

Answers to Checkpoints

Checkpoint 14.1. Sinking depends on pressure, $P = F/A$, not on weight alone, so you would need each one's weight and the total area of the feet or soles in contact with the ground. With the numbers given, the elephant's pressure is larger by $60/20 = 3$: its weight is 60 times larger, but its contact area is only 20 times larger, so the weight ratio wins. The elephant presses on the ground about three times harder per square meter and is the more likely to sink in. (A woman in stiletto heels would reverse the result: see the example "Pressure under a heel," where the heel's tiny area gives a pressure far above an elephant's.)

This page titled 14.1: Fluids, Density, and Pressure is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.2: Pressure as a Function of Depth

Dive to the bottom of a pool and your ears feel the increase at once: pressure grows with depth. The reason is the weight of the fluid overhead, and a free-body diagram makes it exact. Picture an imaginary cylinder of fluid of cross-sectional area A, reaching from the surface (where the pressure is P_0) straight down to a depth h. That cylinder is at rest, so the vertical forces on it must balance. The fluid beneath pushes up on its base with force PA; the air and fluid above push down on its top with P_0A; and gravity pulls down on the fluid inside it with weight $\rho(Ah)g$. Setting the upward force equal to the downward ones,

$$PA = P_0A + \rho Ahg, \tag{14.2.1}$$

and dividing through by A gives

Key Equation: Pressure with depth

$$P = P_0 + \rho gh. \tag{14.3}$$

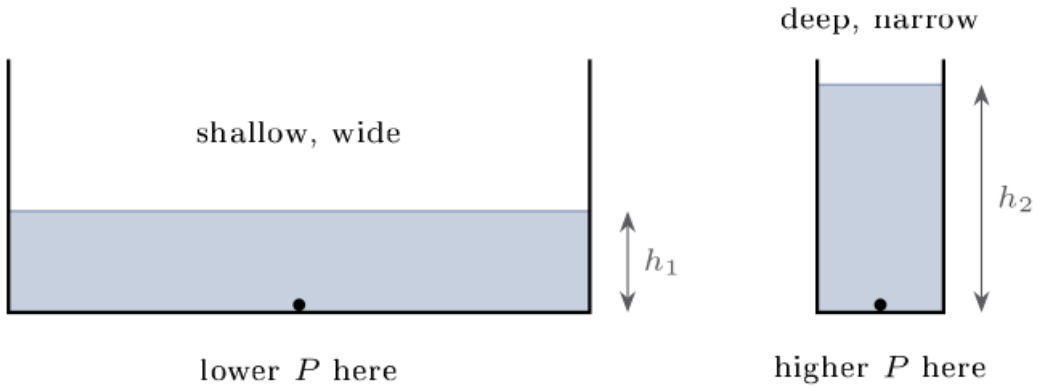


Figure 14.1. Pressure in a fluid depends on depth, not volume. The pressure at the bottom of the deep narrow column (h_2) is greater than at the bottom of the shallow wide one (h_1), even though the wide tank holds roughly twice as much water.

•Figure description

Two connected fluid columns of different shape. The shallow wide column has a smaller depth h sub 1 and lower pressure at its base, while the deep narrow column has a greater depth h sub 2 and higher pressure at its base. Pressure depends on depth, not on the volume of fluid.

Notice what dropped out. The area A canceled, so the result holds for a column of any width, and nothing in the derivation referred to the shape of the container. Pressure depends only on the vertical *depth* below the surface, not on the shape or total volume of the fluid (Figure 14.1): the pressure at the bottom of a narrow deep pond exceeds that at the bottom of a broad shallow lake, even though the lake holds far more water.

Two consequences deserve to be stated on their own, because nearly every fluid-statics problem leans on one of them.

- **Equal depths, equal pressures.** Any two points at the same height within a single *connected* body of fluid at rest are at the same pressure. If they were not, the fluid between them would be pushed from the higher pressure toward the lower, and then it would not be at rest. This is what lets us carry a pressure around a corner: down one arm of a U-tube and up the other, or from the small piston of a hydraulic lift across to the large one.
- **Only vertical distance counts.** In ρgh, the quantity h is the vertical drop, not the distance traveled through the fluid. The water at the bottom of a $10\ \mathrm{m}$ vertical pipe and the water at the bottom of a $50\ \mathrm{m}$ pipe that slopes down through the same $10\ \mathrm{m}$ of elevation are at exactly the same pressure.

The pressure that equation (14.3) returns is the **absolute pressure**, measured up from a perfect vacuum. Most instruments do not report it. A tire gauge, a blood-pressure cuff, and the dial on a scuba tank all measure the amount by which the fluid's pressure exceeds the surrounding atmosphere, called the **gauge pressure**:

$$P_{\text{gauge}} = P_{\text{abs}} - P_{\text{atm}}. \tag{14.4}$$

A flat tire reads zero on a gauge but still holds air at about $1\ \text{atm}$ absolute. For a liquid whose surface is open to the air, $P_0 = P_{\text{atm}}$, and the gauge pressure at depth h is exactly $\rho g h$. Watch for which one a problem wants: "find the pressure at the bottom of the pool" is ambiguous, and the two answers differ by a full atmosphere.

Example 14.2: Water pressure in a house

The main water line enters a house on the first floor at a gauge pressure of $2.86 \times 10^5\ \text{Pa}$. (a) A faucet on the second floor sits $7.10\ \text{m}$ above the line and is turned off. What is the gauge pressure there? (b) How high could a faucet be before no water would reach it at all?

Solution

Both parts use equation (14.3) with one adjustment: here the *known* pressure is at the bottom of the column and we are moving *up*, so the $\rho g h$ term is subtracted rather than added. To keep that straight, write P_{line} for the pressure at the main and P_{faucet} for the pressure a height h above it; do not reuse P_0, which in equation (14.3) is the pressure at the *top*. Both are gauge pressures measured against the same atmosphere, so we may work in gauge values throughout.

(a) The higher faucet has less fluid weight above it, or, equivalently, the line must push water up $7.10\ \text{m}$ to reach it. First compute the pressure of a $7.10\ \text{m}$ water column:

$$\rho g h = (1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)(7.10\,\text{m}) = 6.96 \times 10^4\,\text{Pa}. \tag{14.2.2}$$

(The units: $\text{kg/m}^3 \cdot\ \text{m/s}^2 \cdot\ \text{m} =\ \text{kg/(m}\cdot\text{s}^2) =\ \text{N/m}^2 =\ \text{Pa}$.) Subtract it from the line pressure, writing both over the same power of ten:

$$P_{\text{faucet}} = P_{\text{line}} - \rho g h = 2.86 \times 10^5 - 0.696 \times 10^5 \approx \boxed{2.16 \times 10^5\,\text{Pa}}. \tag{14.2.3}$$

(b) Water just fails to reach a faucet when the gauge pressure there falls to zero. Set $P_{\text{line}} - \rho g h_{\max} = 0$, add $\rho g h_{\max}$ to both sides, and divide by ρg:

$$h_{\max} = \frac{P_{\text{line}}}{\rho g} = \frac{2.86 \times 10^5\,\text{Pa}}{(1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)} = \frac{2.86 \times 10^5}{9800}\,\text{m} \approx \boxed{29.2\,\text{m}}. \tag{14.2.4}$$

No faucet higher than about $29\ \text{m}$ (roughly a ten-story building) could be fed by this line alone. One caution about that number: it is a *static* result, telling you how high the water would stand with every tap closed. Water that is actually flowing loses further pressure to friction in the pipes, so the practical limit is lower. This is why tall buildings need booster pumps or rooftop tanks.

Equation (14.3) also tells us the force on a submerged *surface*, but that force is not simply pressure times area, because the pressure is not the same everywhere on the surface. It grows linearly with depth, so we integrate.

Example 14.3: Force on a dam

A dam holds fresh water to a depth $H = 12\ \text{m}$ against a vertical face of width $w = 30\ \text{m}$. Find the total force the water exerts on the dam.

Solution

Step 1: Why an integral? If the pressure were the same everywhere on the face we could write $F = PA$. It is not: pressure grows with depth. So divide the wetted face into horizontal strips, each thin enough that the pressure across it is essentially uniform. A strip at depth y below the surface, of thickness dy, has area $dA = w\ dy$ and feels a gauge pressure $\rho g y$. The force on that strip is

$$dF = P\,dA = (\rho g y)(w\,dy). \tag{14.2.5}$$

Atmospheric pressure is omitted because it also acts on the dry side of the dam and cancels.

Step 2: Add up the strips. The total force is the integral from the surface ($y = 0$) down to the base ($y = H$). The constants ρ, g, and w come outside, and by the power rule $\int y\, dy = y^2/2$:

$$F = \int_0^H \rho g w\, y\, dy = \rho g w \left[\frac{y^2}{2}\right]_0^H = \rho g w \left(\frac{H^2}{2} - 0\right) = \tfrac{1}{2}\rho g w H^2. \tag{14.2.6}$$

Step 3: Substitute. Square the depth first:

$$\begin{aligned} F &= \tfrac{1}{2}(1000\,\mathrm{kg/m^3})(9.8\,\mathrm{m/s^2})(30\,\mathrm{m})(12\,\mathrm{m})^2 \\ &= \tfrac{1}{2}(1000)(9.8)(30)(144)\,\mathrm{N} \approx \boxed{2.1 \times 10^7\,\mathrm{N}}. \end{aligned} \tag{14.2.7}$$

That is about 2100 tonnes of force.

Check. Rewrite the result as $\left(\rho g \tfrac{H}{2}\right)(wH)$: the pressure at *mid-depth* times the total area. Numerically, $\rho g(H/2) = (1000)(9.8)(6.0) = 5.9 \times 10^4$ Pa and $wH = 360$ m^2, giving 2.1×10^7 N. ✓ The pressure rises linearly from zero, and the average of a linear distribution is its midpoint value. Note also the H^2: doubling the depth quadruples the force, which is why dam walls are built thick at the bottom and thin at the top.

Designing the dam takes one more piece of information: *where* that resultant acts. Because the pressure loading is triangular, the resultant acts one third of the way up from the base, at $H/3$, a torque calculation of the kind you met in Chapter 12.

We live at the bottom of an ocean of air, so we too feel a fluid's weight. The atmosphere, whose bulk lies within about 100 km of the ground, presses on us with an **atmospheric pressure** of about 1.0×10^5 Pa (≈ 15 psi, or 750 mmHg). We are not crushed because our bodies maintain an equal internal pressure that pushes back.

Note: Do not use $P = P_0 + \rho g h$ on the atmosphere

The derivation above assumed the density ρ is the same everywhere in the column. That is excellent for a liquid, which is nearly incompressible: seawater at the bottom of the deepest trench is only about 5% denser than at the surface. It fails badly for a gas, which compresses under its own weight. Air at sea level has $\rho = 1.29$ kg/m^3, but at an altitude of 10 km it is about a third of that. Applying the formula to a 100 km column of sea-level air would give $\rho g h = (1.29)(9.8)(1.0 \times 10^5) \approx 1.3 \times 10^6$ Pa, roughly 13 atmospheres: more than ten times the true value. Real atmospheric pressure falls off approximately exponentially, dropping by a factor of e about every 8.5 km.

Checkpoint 14.2

A student objects: "The pressure at the bottom of a 10 m pipe of 2 cm diameter can't be as large as at the bottom of a 10 m pool; the pipe holds only a few liters of water, and the pool holds hundreds of tonnes." Identify the confusion, and use the free-body derivation to say precisely what the extra water in the pool is being supported by, if not by the water at the bottom. Then answer this: a reservoir behind a dam is 40 m deep and 5 km long; a second reservoir is 40 m deep and 200 m long. Do their dams need to be built to the same strength? *(Careful: "strength" involves both pressure and total force.)*

Answer

Answers to Checkpoints

Checkpoint 14.2. The confusion is between pressure and total weight. The pressure at depth h is $\rho g h$ however wide the container is: the free-body derivation balances the weight of a column of water directly *above* a small area at the bottom, and a column 10 m tall has the same weight per unit area whether it stands in a pipe or in a pool. The rest of the pool's water is not pressing on that spot at all; it rests on the rest of the pool floor, which supports it. The two dams need the *same* strength per meter of their width: the pressure at every depth, and hence the force on each strip of the dam's face, depends only on the depth (40 m in both), not on how far the water extends behind the dam. Total force does depend on the dam's *width* across the valley, $F = \tfrac{1}{2}\rho g w H^2$, but the reservoir's length appears nowhere.

This page titled 14.2: Pressure as a Function of Depth is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.3: Pascal's Principle

What happens if we squeeze a confined fluid? Because a fluid flows, any increase in pressure at one point is passed along to every other point. This is **Pascal's principle**: a pressure change applied to an enclosed fluid is transmitted undiminished to every part of the fluid and to the walls of its container.

The principle is the basis of all **hydraulic** machines: car brakes, lifts, and presses. Apply a force F_1 to a small piston of area A_1, and the pressure $P = F_1/A_1$ appears throughout the fluid, including at a large piston of area A_2, where it produces a force $F_2 = PA_2$. Equating the pressures gives

Key Equation: Pascal's principle (hydraulics)

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}. \tag{14.5}$$

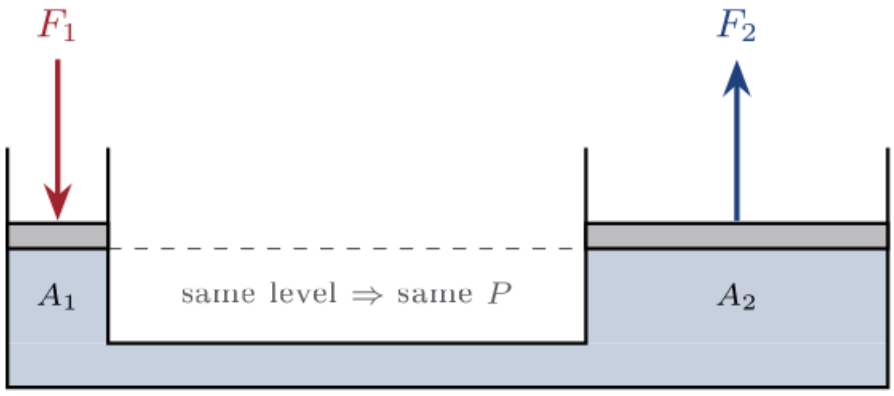


Figure 14.2. A hydraulic lift. A small force F_1 on the narrow piston creates a pressure transmitted through the connected fluid to the wide piston, where it produces a much larger force $F_2 = F_1(A_2/A_1)$. The two pistons are drawn at the same level, so the pressures at their faces are equal.

•Figure description

A hydraulic lift. A small force F sub 1 acts on a narrow piston of area A sub 1, and the pressure transmitted through the connected fluid produces a much larger force F sub 2 on a wide piston of area A sub 2.

A small force on a small piston balances a large force on a large piston, in proportion to the areas: a multiplication of force, the hydraulic version of the lever (Figure 14.2).

Example 14.4: A hydraulic lift

The small piston of a hydraulic lift has area $5.0\ \text{cm}^2$; the large piston has area $250\ \text{cm}^2$. What force on the small piston is needed to support a $5000\ \text{N}$ car on the large one?

Solution

By Pascal's principle, the pressure is the same under both pistons, $P_1 = P_2$. Writing each pressure as force over area (equation (14.5)),

$$\frac{F_1}{A_1} = \frac{F_2}{A_2}. \tag{14.3.1}$$

Multiply both sides by A_1 to isolate F_1:

$$F_1 = F_2\,\frac{A_1}{A_2} = (5000\,\mathrm{N})\,\frac{5.0\ \cancel{\mathrm{cm}^2}}{250\ \cancel{\mathrm{cm}^2}} = (5000\,\mathrm{N})(0.020) = \boxed{100\,\mathrm{N}}. \tag{14.3.2}$$

Only the *ratio* of the areas matters, so the square centimeters cancel and there is no need to convert to square meters.

A $100\ \mathrm{N}$ push (about the weight of a $10\ \mathrm{kg}$ sack of flour) holds up a car, because the large piston has $250/5.0 = 50$ times the area. (The force is multiplied, but the small piston must move 50 times as far: the fluid volume pushed out of the small cylinder, $A_1 d_1$, must equal the volume pushed into the large one, $A_2 d_2$. The work in, $F_1 d_1$, therefore equals the work out, $F_2 d_2$: energy is conserved.)

Note: The two pistons should be at the same height

Equation (14.5) sets the pressures at the two piston faces equal, which is exactly true only when they are at the same level. If the output piston sits a height d above the input piston, the pressure there is lower by $\rho g d$, and $F_2 = (F_1/A_1 - \rho g d)\,A_2$. For a garage lift with $d \approx 1.5\ \mathrm{m}$ of hydraulic oil ($\rho \approx 850\ \mathrm{kg/m^3}$), that correction is about $1.2 \times 10^4\ \mathrm{Pa}$ against working pressures of $1 \times 10^7\ \mathrm{Pa}$ or more: roughly 0.1%, which is why it is normally ignored. In a tall system it is not: the water column feeding a high-rise, or a hydraulic line running several stories, has to be designed around it.

Checkpoint 14.3

A hydraulic lift's input piston has $1/50$ the area of its output piston. A student proposes shrinking the input piston until the ratio is $1/100$, so that the same push lifts twice the load, and asks why lifts are not simply built this way. Say what the change costs (be quantitative about the distance the operator's piston must travel) and identify a practical limit that has nothing to do with energy. Then explain a related point: hydraulic systems use oil rather than air, and mechanics "bleed" brake lines to chase out every bubble. What goes wrong if a little air is trapped in the line, given that Pascal's principle applies to gases too? *(Hint: the principle transmits the pressure faithfully; what it does not promise is that the fluid keeps its volume.)*

Answer

Answers to Checkpoints

Checkpoint 14.3. With a $1/100$ area ratio the output force doubles, but the fluid volume is unchanged, so the input piston must travel 100 times as far as the load rises, instead of 50. The work in still equals the work out; the operator simply pushes half as hard for twice the distance. The practical limit is that stroke: lifting a car $2\ \mathrm{m}$ would take $200\ \mathrm{m}$ of piston travel, meaning an absurdly long cylinder or a great many pump strokes, with seals and leakage losses on every one. Trapped air ruins the system because air is compressible. Pascal's principle still transmits the pressure, but when you push the pedal the bubble shrinks instead of the liquid moving. The pedal travels without moving the brake pads until the air is squeezed down, giving the familiar "spongy" brakes. Liquids are nearly incompressible, so a push on one end moves the other end at once.

This page titled 14.3: Pascal's Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.4: Buoyancy and Archimedes' Principle

Why does a heavy object feel lighter underwater? Because pressure increases with depth, the fluid pushes up on the bottom of a submerged object harder than it pushes down on the top. The net upward force is the **buoyant force**.

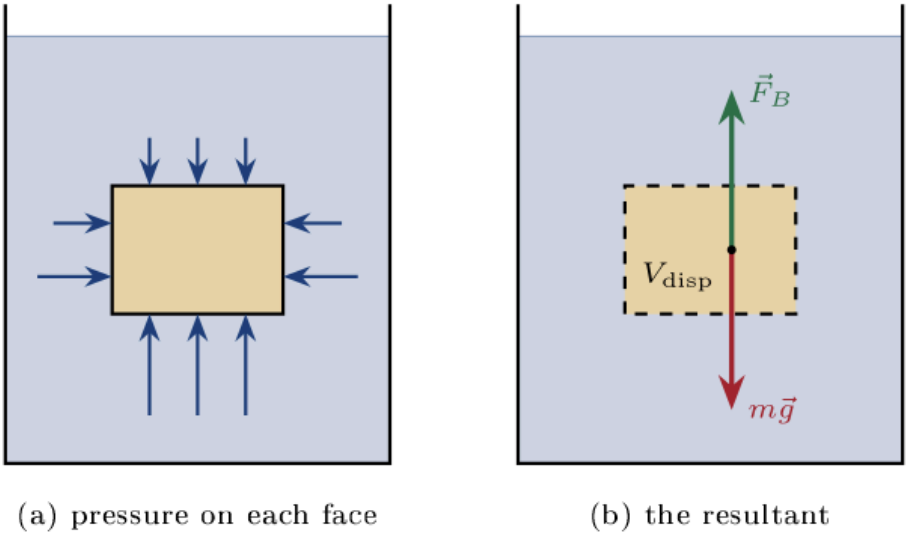


Figure 14.3. Buoyancy. (a) The fluid presses inward on every face. The side pushes are at equal depths and cancel; the push on the base is larger than the push on the top because the base is deeper. (b) The net result is a single upward buoyant force equal to the weight of the fluid that would fill the dashed volume. Here F_B and mg happen to balance, so the block hovers; if they do not, it rises or sinks.

•Figure description

Two panels on buoyancy. In panel (a) the fluid presses inward on every face of a submerged block, with arrows that start in the fluid and end on each face, longer at greater depth; the sideways pushes are at equal depths and cancel, while the upward push on the base exceeds the downward push on the top. Panel (b) shows the resultant, a net upward buoyant force equal to the weight of the displaced fluid.

Its size is easy to pin down for a rectangular block of horizontal face area A, held with its top face at depth h_1 and its bottom face at depth h_2. The fluid presses inward on all six faces. The pushes on opposite vertical sides are at equal depths, so they cancel; the vertical ones do not, because the bottom sits deeper:

$$F_B = P_{\text{bottom}}A - P_{\text{top}}A = \big[(P_0 + \rho g h_2) - (P_0 + \rho g h_1)\big]A = \rho g(h_2 - h_1)A = \rho g V, \tag{14.4.1}$$

where $V = (h_2 - h_1)A$ is the block's volume, precisely the volume of fluid it pushed out of the way. Remarkably, the shape turns out not to matter at all, a result known as **Archimedes' principle**: a body immersed in a fluid is buoyed up by a force equal to the weight of the fluid it displaces.

Key Equation: Archimedes' principle

$$F_B = \rho_{\text{fluid}}\, V_{\text{disp}}\, g, \tag{14.6}$$

where V_{disp} is the volume of fluid displaced (Figure 14.3). For a fully submerged object this is the object's whole volume; for a floating one it is only the part below the surface.

Whether an object sinks or floats follows from comparing equation (14.6) with the object's weight. Submerge an object of volume V and average density ρ_{obj}: the buoyant force is $\rho_{\text{fluid}} V g$ and the weight is $\rho_{\text{obj}} V g$. The volume cancels, so only the densities matter:

$$\rho_{\text{obj}} > \rho_{\text{fluid}} \;\Rightarrow\; \text{sinks}, \qquad \rho_{\text{obj}} < \rho_{\text{fluid}} \;\Rightarrow\; \text{rises}, \qquad \rho_{\text{obj}} = \rho_{\text{fluid}} \;\Rightarrow\; \text{hovers}. \tag{14.4.2}$$

The third case, in which an object hangs motionless at any depth, is called **neutral buoyancy**; it is what a scuba diver trims for.

It is the *average* density that counts, which is how a steel ship floats: hollow and full of air, the hull-plus-air combination is less dense than water, even though steel alone is eight times denser. Submarines exploit the same idea, flooding or emptying tanks to shift their average density above or below that of seawater.

An object that rises does not leave the fluid entirely; it settles at the depth where it displaces exactly its own weight of fluid. Setting weight equal to buoyant force for a floating body, $\rho_{\text{obj}} V g = \rho_{\text{fluid}} V_{\text{sub}}\ g$, gives the submerged fraction directly:

Key Equation: Floating: fraction submerged

$$\frac{V_{\text{sub}}}{V} = \frac{\rho_{\text{obj}}}{\rho_{\text{fluid}}}. \tag{14.7}$$

Ice ($917\ \text{kg/m}^3$) in seawater ($1025\ \text{kg/m}^3$) floats with 89% of its volume submerged, hence the iceberg's hidden nine-tenths. Only the ratio of densities appears, which is why the same log floats a little lower in fresh water than in salt water, and why a ship rides deeper leaving the ocean for a river.

Example 14.5: Raising a sunken container

A sealed, watertight shipping container resting on the ocean floor measures $6.2\ \text{m} \times 2.7\ \text{m} \times 2.3\ \text{m}$. Salvagers attach a spherical balloon to its top and inflate it; when the balloon's radius reaches $1.4\ \text{m}$, the container just begins to rise. Find the mass of the container. The density of seawater is $1025\ \text{kg/m}^3$; ignore the mass of the balloon and the air inside it, and ignore any suction from the mud.

Solution

Step 1: The condition. The container "just begins to rise" when the total buoyant force equals the container's weight. Both the container and the balloon displace seawater, so the displaced volume is the sum of the two.

Step 2: The volumes. The container is a box, length times width times height; the balloon is a sphere, $\frac{4}{3}\pi r^3$, with the radius cubed before multiplying:

$$\begin{aligned} V_{\text{box}} &= (6.2)(2.7)(2.3)\,\text{m}^3 = 38.5\,\text{m}^3, \\ V_{\text{balloon}} &= \tfrac{4}{3}\pi(1.4\,\text{m})^3 = \tfrac{4}{3}\pi(2.744\,\text{m}^3) = 11.5\,\text{m}^3, \end{aligned} \tag{14.4.3}$$

giving $V_{\text{disp}} = 38.5 + 11.5 = 50.0\ \text{m}^3$.

Step 3: Solve for the mass. By Archimedes' principle the buoyant force is the weight of the displaced water, $\rho_{\text{sw}} V_{\text{disp}}\ g$. Set it equal to the container's weight mg (the balloon and air masses are negligible), and divide both sides by g:

$$\rho_{\text{sw}}\, V_{\text{disp}}\, g = mg \implies m = \rho_{\text{sw}} V_{\text{disp}} = (1025\,\text{kg/m}^3)(50.0\,\text{m}^3) \approx \boxed{5.1 \times 10^4\,\text{kg}}. \tag{14.4.4}$$

The value of g drops out entirely: the same salvage would work on any planet with seawater.

The container has a mass of about 51 tonnes. Squeezing that into a box of $38.5\ \text{m}^3$ means dense cargo (machinery or metal stock), and it is worth seeing why the problem had to be set up that way. A container loaded to its usual legal maximum of about 30 tonnes has an average density of $\rho = m/V = (30\ 000\ \text{kg})/(38.5\ \text{m}^3) \approx 780\ \text{kg/m}^3$, well below that of seawater. Such a container does not sink at all: containers lost overboard often drift half-submerged for weeks, a genuine hazard to small craft.

Checkpoint 14.4

An ice cube floats in a glass of water filled exactly to the brim. As it melts, does the glass overflow? Answer using equation (14.7) rather than intuition, and then say what changes if the ice cube has a small steel bolt frozen inside it. Now a harder version of the same idea: you are sitting in a small boat in a pond, holding a heavy rock. You throw the rock overboard and it sinks. Does the water level of the pond rise, fall, or stay the same? *(Hint: compare the volume of water the rock displaces sitting on the bottom with the volume it displaced while riding in the boat.)*

Answer

Answers to Checkpoints

Checkpoint 14.4. No. By equation (14.7), a floating cube displaces its own *weight* of water. When it melts it becomes water of that same weight, which exactly fills the volume it was already displacing, so the level does not change. With a steel bolt inside, the floating cube displaces the weight of ice *plus* bolt. After melting, the bolt sinks and displaces only its own *volume*, which is much less than the volume of water equal to its weight (steel is about 8 times denser than water). The level falls, so there is still no overflow. The rock works the same way: in the boat it displaces its weight of water; on the bottom it displaces only its volume, which is smaller. So the pond's water level *falls*.

This page titled 14.4: Buoyancy and Archimedes' Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.5: Fluid Dynamics and Bernoulli's Principle

So far the fluid has been at rest. When it flows, we idealize it as an **ideal fluid** (incompressible, nonviscous, and in steady flow), which captures the essential physics without the complications of turbulence and friction.

The continuity equation

Consider fluid flowing through a pipe that narrows. Since fluid is neither created nor destroyed, the mass passing any cross-section per second must be the same everywhere: what goes in must come out. The mass flow rate is $\rho A v$ (density times area times speed), so $\rho_1 A_1 v_1 = \rho_2 A_2 v_2$. For an incompressible fluid the density is constant, leaving the **equation of continuity**:

Key Equation: Equation of continuity

$$A_1 v_1 = A_2 v_2. \tag{14.8}$$

The product Av that appears on both sides has a name of its own. It is the **volume flow rate**,

$$Q = Av, \tag{14.9}$$

measured in $\mathrm{m^3/s}$: the volume of fluid crossing a section each second. Continuity says simply that Q is the same at every point along the pipe. Multiplying by density gives the **mass flow rate** $\rho A v = \rho Q$, in $\mathrm{kg/s}$, which is how pumps and compressors are usually rated. Beware of units here: flow rates are quoted in liters per second, gallons per minute, and cubic feet per hour, and all of them must be converted to $\mathrm{m^3/s}$ before use ($1\ \mathrm{L} = 1 \times 10^{-3}\ \mathrm{m^3}$).

Where the pipe is narrow, the fluid speeds up; where it widens, it slows down. This is why a partly covered hose nozzle shoots water faster and farther.

Example 14.6: Gasoline in a fuel line

A fuel pump delivers gasoline ($\rho = 734\ \mathrm{kg/m^3}$) at a mass flow rate of $6.22 \times 10^{-2}\ \mathrm{kg/s}$ through a fuel line of radius $2.34 \times 10^{-3}\ \mathrm{m}$. How fast does the gasoline move along the line?

Solution

Step 1: Rearrange. The mass flow rate is the mass of fluid passing a point each second, $\rho A v$ (density times volume flow rate). We want v, so divide both sides by ρA:

$$\frac{\Delta m}{\Delta t} = \rho A v \quad \Longrightarrow \quad v = \frac{\Delta m/\Delta t}{\rho A}. \tag{14.5.1}$$

Step 2: The area. The line is circular, so $A = \pi r^2$. Squaring 2.34×10^{-3} squares both parts: $(2.34)^2 = 5.48$ and $(10^{-3})^2 = 10^{-6}$. So

$$A = \pi(2.34 \times 10^{-3}\,\mathrm{m})^2 = \pi(5.48 \times 10^{-6})\,\mathrm{m^2} = 1.72 \times 10^{-5}\,\mathrm{m^2}. \tag{14.5.2}$$

Step 3: Substitute.

$$v = \frac{6.22 \times 10^{-2}\,\mathrm{kg/s}}{(734\,\mathrm{kg/m^3})(1.72 \times 10^{-5}\,\mathrm{m^2})} = \frac{6.22 \times 10^{-2}}{1.26 \times 10^{-2}}\,\mathrm{m/s} \approx \boxed{4.93\,\mathrm{m/s}}. \tag{14.5.3}$$

The units reduce correctly: $(\ \mathrm{kg/s})/(\ \mathrm{kg/m^3} \cdot\ \mathrm{m^2}) = (\ \mathrm{kg/s})/(\ \mathrm{kg/m}) =\ \mathrm{m/s}$.

Example 14.7: How far does the hose squirt?

A gardener fills a $30.0\ \mathrm{L}$ bucket in $1.00\ \mathrm{min}$, then attaches a nozzle of opening area $0.500\ \mathrm{cm}^2$ and holds it horizontally $1.00\ \mathrm{m}$ above the ground. Over what horizontal distance does the water land? Neglect air resistance.

Solution

Step 1: Volume flow rate in SI. A liter is $10^{-3}\ \mathrm{m}^3$, so $30.0\ \mathrm{L} = 3.00 \times 10^{-2}\ \mathrm{m}^3$, and a minute is $60.0\ \mathrm{s}$:

$$Q = \frac{3.00 \times 10^{-2}\,\mathrm{m}^3}{60.0\,\mathrm{s}} = 5.00 \times 10^{-4}\,\mathrm{m}^3/\mathrm{s}. \tag{14.5.4}$$

Step 2: Speed at the nozzle. The hose's flow rate does not change when the nozzle is attached; the water is simply forced through a smaller opening. Convert the area with the squared factor, $1\ \mathrm{cm}^2 = 10^{-4}\ \mathrm{m}^2$, so $A = 0.500 \times 10^{-4}\ \mathrm{m}^2 = 5.00 \times 10^{-5}\ \mathrm{m}^2$. Then $Q = Av$ gives, on dividing by A,

$$v = \frac{Q}{A} = \frac{5.00 \times 10^{-4}\,\mathrm{m}^3/\mathrm{s}}{5.00 \times 10^{-5}\,\mathrm{m}^2} = 10.0\,\mathrm{m}/\mathrm{s}. \tag{14.5.5}$$

Step 3: Projectile motion. Once the water leaves the nozzle it is a projectile (Chapter 4), launched horizontally at $10.0\ \mathrm{m}/\mathrm{s}$. As in Chapter 4, the vertical motion sets the time: starting with no vertical velocity, it falls $h = \frac{1}{2}gt^2$. Multiply by $2/g$ and take the square root:

$$t = \sqrt{\frac{2h}{g}} = \sqrt{\frac{2(1.00\,\mathrm{m})}{9.8\,\mathrm{m}/\mathrm{s}^2}} = \sqrt{0.204\,\mathrm{s}^2} = 0.452\,\mathrm{s}. \tag{14.5.6}$$

In that time the water travels horizontally at its constant nozzle speed, so the range is

$$x = v\,t = (10.0\,\mathrm{m}/\mathrm{s})(0.452\,\mathrm{s}) \approx \boxed{4.52\,\mathrm{m}}. \tag{14.5.7}$$

Checkpoint 14.5

Turn on a faucet gently and watch the falling stream: it grows visibly narrower as it descends, sometimes breaking into drops. Explain the narrowing using equation (14.8), being explicit about what makes the water speed up. Then decide whether the following claim is right: "Continuity says the water speeds up in the narrow part of a pipe, so it must gain kinetic energy, so continuity violates conservation of energy." Identify the error, and name what supplies the energy.

Answer

Bernoulli's equation

Continuity tells us a fluid speeds up in a constriction, but speeding up requires a net force, which means the pressure must be *lower* where the fluid moves faster. This is **Bernoulli's principle**: where the speed of a fluid increases, its pressure decreases.

The quantitative statement follows from the energy bookkeeping of Chapter 8 (equation (8.1)) applied to a parcel of fluid and the Earth: the work done on the parcel by pressure differences equals the change in its kinetic plus gravitational potential energy; dividing through by the volume and writing $\rho = m/V$ yields **Bernoulli's equation**:

Key Equation: Bernoulli's equation

$$P + \tfrac{1}{2}\rho v^2 + \rho g y = \text{constant}. \tag{14.10}$$

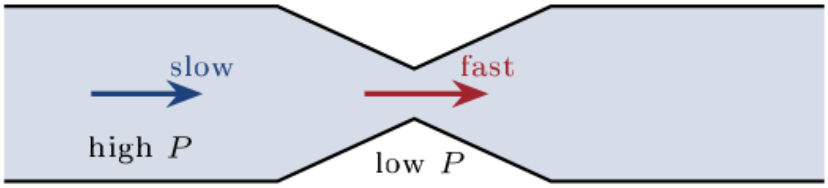


Figure 14.4. Bernoulli's principle in a constricted pipe. By continuity the fluid moves faster through the narrow section; by Bernoulli's equation the pressure there is lower.

•Figure description

A pipe that narrows in the middle. The fluid moves slowly through the wide section where the pressure is high, and faster through the narrow section where the pressure is low.

Each term is an energy per unit volume: pressure energy, kinetic energy, and gravitational potential energy. Along a streamline their sum is conserved. Bernoulli's equation explains a host of everyday effects: the lift on an airplane wing (air moves faster over the curved top, lowering the pressure there),[1] the curve of a spinning baseball, and the ventilation of an animal's burrow. Figure 14.4 shows the pressure–speed trade-off in its simplest form, a constricted pipe.

Example 14.8: Ventilating a prairie-dog burrow

A prairie-dog burrow has two openings. Opening 1 is a plain hole at ground level; opening 2 sits atop a raised mound about $0.20\ \mathrm{m}$ higher, where the wind runs clear of the ground and therefore faster. Wind blows across opening 1 at $8.5\ \mathrm{m/s}$ and across opening 2 at $12\ \mathrm{m/s}$. Taking the density of air to be $1.29\ \mathrm{kg/m^3}$, find the pressure difference $\Delta P = P_1 - P_2$ that drives air through the burrow.

Solution

Step 1: Bernoulli between the two openings. Both airstreams come from the same undisturbed air far upwind, so they share the same Bernoulli constant and we may write

$$P_1 + \tfrac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \tfrac{1}{2}\rho v_2^2 + \rho g y_2. \tag{14.5.8}$$

To find $\Delta P = P_1 - P_2$, subtract P_2 from both sides and move every term without a P to the right; then factor $\tfrac{1}{2}\rho$ out of the speed terms and ρg out of the height terms:

$$\Delta P = P_1 - P_2 = \tfrac{1}{2}\rho\left(v_2^2 - v_1^2\right) + \rho g\,(y_2 - y_1). \tag{14.5.9}$$

Step 2: Is the height term worth keeping? Check it before carrying it: $\rho g(y_2 - y_1) = (1.29)(9.8)(0.20) = 2.5\ \mathrm{Pa}$, about 5% of the speed term below, smaller than the uncertainty in wind speeds quoted to two figures. We drop it. (This is typical of gases: because air is so light, $\rho g y$ matters only over height changes of tens of meters.)

Step 3: The speed term. Square each speed *before* subtracting, $12^2 = 144$ and $8.5^2 = 72.25$:

$$\Delta P = \tfrac{1}{2}(1.29\,\mathrm{kg/m^3})\left(144 - 72.25\right)\mathrm{m^2/s^2} = \tfrac{1}{2}(1.29)(71.75)\,\mathrm{Pa} \approx \boxed{46\,\mathrm{Pa}}. \tag{14.5.10}$$

Subtracting the speeds first, $(12 - 8.5)^2 = 12.25$, would give only $8\ \mathrm{Pa}$, the same slip as in Chapter 8, and for the same reason: pressure, like kinetic energy, goes as the square of the speed.

Since $\Delta P > 0$, the pressure is higher at the low, slow opening. Air enters there, sweeps through the burrow, and leaves at the mound: natural ventilation, courtesy of Bernoulli. Prairie dogs really do build one entrance mound higher than the other, which is exactly what creates the difference in wind speed.

Note: Where Bernoulli's equation stops working

Every term in equation (14.10) is an energy per unit volume, and the equation asserts that none of that energy is lost. Real fluids lose it steadily to viscosity, the internal friction that our "ideal fluid" assumption threw away. Across a short constriction or a nozzle the loss is negligible and Bernoulli is excellent. Along $100\ \mathrm{m}$ of pipe it dominates: moving water through a pipeline

is almost entirely a fight against friction, and engineers handle it with empirical relations such as the Darcy–Weisbach equation rather than with Bernoulli. The equation also fails outright across anything that adds or removes energy (a pump, fan, or turbine) and downstream of anything that churns the flow into turbulence. Use it for short, smooth, steady runs; be suspicious everywhere else.

Checkpoint 14.6

The fabric roof of a convertible bulges *upward* at highway speed, and the same thing happens to the roof of a house in a strong wind: roofs are more often lifted off than pushed in. Explain both using equation (14.10), stating clearly which pressures you are comparing and where the air inside is moving. Then estimate: for a $30\ \mathrm{m/s}$ wind blowing across a flat roof of area $150\ \mathrm{m^2}$, roughly what upward force does the pressure difference produce? Compare it with the weight of the roof structure, and say whether you trust the estimate.

Answer

---

[1] Bernoulli's equation correctly links the faster flow over the top of a wing to the lower pressure there, but it does not by itself explain *why* the air moves faster. The popular story (that two air parcels separated at the leading edge must meet again at the trailing edge, so the one taking the longer path over the top must hurry) is simply false; the air crossing the top arrives well ahead of the air underneath. A full account rests on the circulation established by the wing's shape and angle of attack, or equivalently on the fact that the wing deflects a great deal of air downward and gets pushed up in reaction (Newton's third law). You will meet the complete story in a course on fluid mechanics or aerodynamics. ↩

Answers to Checkpoints

Checkpoint 14.5. Falling water speeds up under gravity. For water, which is essentially incompressible, the volume flow rate Av is the same at every height (equation (14.8)), so as v grows the cross-sectional area A must shrink; the stream narrows. The claim is wrong because continuity says nothing about where the energy comes from. In a pipe, the water entering the narrow section is pushed by higher pressure behind it than ahead of it, and that pressure difference does the work that supplies the extra kinetic energy. This is exactly Bernoulli's equation: where the speed rises, the pressure falls. In the falling stream, gravity supplies the energy.

Checkpoint 14.6. Outside, air streams rapidly over the roof, so by Bernoulli's equation its pressure there is below atmospheric. Inside the car or house the air is nearly still, so it stays at about atmospheric pressure. The higher pressure inside pushes up on the roof harder than the lower pressure outside pushes down, so the net force is upward. Estimate: $\Delta P \approx \frac{1}{2}\rho v^2 = \frac{1}{2}(1.2)(30)^2 \approx 540\ \mathrm{Pa}$, so $F \approx (540)(150) \approx 8 \times 10^4\ \mathrm{N}$, the weight of about 8 tonnes. That is comparable to or larger than the weight of a typical house roof (20–$50\ \mathrm{kg}$ per square meter, so a few tonnes), which is why roofs must be strapped down. Trust it as an order-of-magnitude estimate only: the real flow over a roof is turbulent and uneven, and open windows change the inside pressure, but the conclusion that wind can lift a roof off holds up.

This page titled 14.5: Fluid Dynamics and Bernoulli's Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.S: Fluids (Summary)

Summary

- A **fluid** flows and takes its container's shape. Its **density** is $\rho = m/V$, and **pressure** is the perpendicular force per area, $P = F/A$ (in pascals). Pressure is a scalar: at a point in a fluid it acts equally in all directions.
- Pressure increases with depth: $P = P_0 + \rho g h$, depending on the vertical depth, not on volume or container shape. In one connected fluid at rest, all points at the same height are at the same pressure. The relation assumes constant density, so it applies to liquids but not to the atmosphere as a whole.
- **Gauge pressure** is the excess over the local atmosphere, $P_{\text{gauge}} = P_{\text{abs}} - P_{\text{atm}}$; for a liquid open to the air it is just $\rho g h$. We live under an atmospheric pressure of about $1 \times 10^5 \text{ Pa} = 1 \text{ atm} = 14.7 \text{ psi}$.
- **Pascal's principle**: a pressure change in an enclosed fluid is transmitted undiminished, so a hydraulic system multiplies force, $F_1/A_1 = F_2/A_2$. Force is multiplied; work is not.
- **Archimedes' principle**: the buoyant force equals the weight of displaced fluid, $F_B = \rho_{\text{fluid}} V_{\text{disp}}\ g$. An object sinks or rises according to whether its *average* density exceeds that of the fluid, and a floating body submerges the fraction $V_{\text{sub}}/V = \rho_{\text{obj}}/\rho_{\text{fluid}}$.
- For an ideal fluid in motion, the **continuity equation** $A_1 v_1 = A_2 v_2$ keeps the volume flow rate $Q = Av$ constant along a pipe, and **Bernoulli's equation** $P + \frac{1}{2}\rho v^2 + \rho g y = \text{constant}$ conserves energy per unit volume, so faster flow means lower pressure. Bernoulli holds only where viscous losses, turbulence, and pumps can be ignored.

Key Equations

Density

$$\rho = \frac{m}{V} \tag{14.S.1}$$

Pressure

$$P = \frac{F}{A} \tag{14.S.2}$$

Pressure with depth

$$P = P_0 + \rho g h \tag{14.S.3}$$

Pascal's principle (hydraulics)

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} \tag{14.S.4}$$

Archimedes' principle

$$F_B = \rho_{\text{fluid}}\, V_{\text{disp}}\, g \tag{14.S.5}$$

Floating: fraction submerged

$$\frac{V_{\text{sub}}}{V} = \frac{\rho_{\text{obj}}}{\rho_{\text{fluid}}} \tag{14.S.6}$$

Equation of continuity

$$A_1 v_1 = A_2 v_2 \tag{14.S.7}$$

Bernoulli's equation

$$P + \tfrac{1}{2}\rho v^2 + \rho g y = \text{constant} \tag{14.S.8}$$

This page titled 14.S: Fluids (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

14.E: Fluids (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging.

Section 14.1 — Fluids, Density, and Pressure

14.1 ★ A block of aluminum has a mass of $2.7\ \text{kg}$ and a volume of $1.0 \times 10^{-3}\ \text{m}^3$. Find its density, and state whether it would sink or float in water.

14.2 ★ *[Conceptual.]* A sharp knife cuts far more easily than a dull one, even with the same force behind it. Explain this using the definition of pressure.

14.3 ★★ *[Real-world: engineering.]* A rectangular water bed measures $2.0\ \text{m} \times 1.5\ \text{m}$ and is $0.30\ \text{m}$ thick.

a. Find the mass of water it holds ($\rho_{\text{water}} = 1000\ \text{kg/m}^3$).
b. Find the pressure it exerts on the floor, assuming it rests on its full $2.0\ \text{m} \times 1.5\ \text{m}$ area.

Section 14.2 — Pressure as a Function of Depth

14.4 ★ Find the gauge pressure at the bottom of a swimming pool $3.0\ \text{m}$ deep.

14.5 ★ *[Conceptual.]* Two open tanks hold water to the same depth, but one is ten times wider than the other. Compare the pressure at the bottom of each, and explain why volume does not matter.

14.6 ★★ *[Real-world: oceanography.]* A submarine dives to a depth of $250\ \text{m}$ in seawater ($\rho = 1025\ \text{kg/m}^3$).

a. Find the gauge pressure at that depth.
b. Find the total (absolute) pressure, including atmospheric pressure ($1.0 \times 10^5\ \text{Pa}$).

14.7 ★★ A dam holds back a lake. At a point $12\ \text{m}$ below the surface, find the gauge pressure of the water against the dam, and the force that pressure exerts on a $1.0\ \text{m}^2$ inspection panel centered at that depth. *(The panel is small enough that the pressure may be treated as uniform across it.)*

Section 14.3 — Pascal's Principle

14.8 ★ In a hydraulic lift, the input piston has area $8.0\ \text{cm}^2$ and the output piston $400\ \text{cm}^2$. What force on the input piston supports a $12000\ \text{N}$ load?

14.9 ★★ *[Real-world: automotive.]* A hydraulic brake system applies a force of $90\ \text{N}$ to a master-cylinder piston of area $1.2\ \text{cm}^2$. Each brake (slave) cylinder has area $6.0\ \text{cm}^2$. Find the force exerted at each brake.

14.10 ★★ *[Conceptual.]* A hydraulic lift multiplies force, letting a small push lift a heavy car. Does it also multiply energy or work? Explain what happens to the distances the two pistons move, and why energy is still conserved.

Section 14.4 — Buoyancy and Archimedes' Principle

14.11 ★ A rock displaces $2.0 \times 10^{-3}\ \text{m}^3$ of water when fully submerged. Find the buoyant force on it.

14.12 ★ *[Conceptual.]* A steel ship floats, yet a solid steel ball of the same mass sinks. Explain the difference in terms of displaced water and Archimedes' principle.

14.13 ★★ A wooden block of density $650\ \text{kg/m}^3$ floats in water. What fraction of its volume is submerged? *(Hint: a floating object displaces its own weight of fluid.)*

14.14 ★★★ A $0.500\ \text{kg}$ crown is suspended from a scale and lowered fully into water; the scale then reads $4.10\ \text{N}$.

a. Find the buoyant force on the crown.
b. Find the crown's volume and density.
c. Is it pure gold ($\rho = 19300\ \text{kg/m}^3$)?

14.15 ★★★ *[Real-world: recreation.]* A $75\ \text{kg}$ swimmer floats in seawater of density $1025\ \text{kg/m}^3$, with 2.0% of her volume above the surface. Find her total body volume and her average density.

Section 14.5 — Fluid Dynamics and Bernoulli's Principle

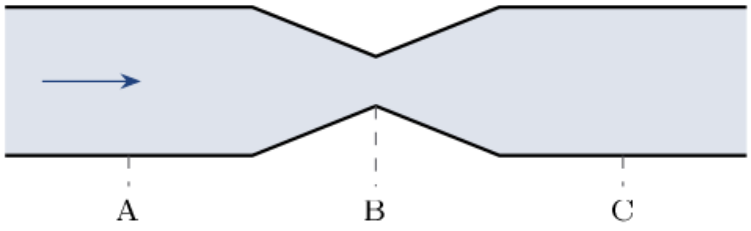
14.16 ★ Water flows through a pipe of cross-sectional area $0.080\ \text{m}^2$ at $2.0\ \text{m/s}$. The pipe narrows to $0.020\ \text{m}^2$. Find the water's speed in the narrow section.

14.17 ★ *[Conceptual.]* Hold two sheets of paper hanging down a few centimeters apart and blow between them. They move *together*, not apart. Explain using Bernoulli's principle.

14.18 ★★ *[Real-world: firefighting.]* Water flows at $1.5\ \text{m/s}$ through a fire hose of radius $3.0\ \text{cm}$, then exits a nozzle of radius $1.0\ \text{cm}$.

a. Find the speed of the water leaving the nozzle.
b. If the nozzle is aimed horizontally $1.5\ \text{m}$ above the ground, how far horizontally from the nozzle does the stream land? Neglect air resistance.

14.19 ★★ *[Diagram-reading.]* Water flows steadily left to right through the horizontal pipe shown, whose cross-section narrows in the middle.



•Figure description
A horizontal pipe carrying water left to right, with its cross-section narrowing in the middle. Three points are marked A in the wide inlet, B in the narrow throat, and C in the wide outlet.

a. Rank the flow speeds at A, B, and C.
b. Rank the pressures at A, B, and C. Explain using continuity and Bernoulli's principle.

14.20 ★★★ Water flows through a horizontal pipe that narrows from a radius of $4.0\ \text{cm}$ to $2.0\ \text{cm}$. In the wide section the speed is $1.2\ \text{m/s}$ and the gauge pressure is $3.0 \times 10^4\ \text{Pa}$. Find (a) the speed and (b) the gauge pressure in the narrow section.

Additional Problems

14.21 ★★★ A U-shaped tube open at both ends contains water ($\rho_w = 1000\ \text{kg/m}^3$). Oil ($\rho_{\text{oil}} = 850\ \text{kg/m}^3$) is slowly poured into the left arm until it forms a column $8.0\ \text{cm}$ tall floating on the water.

a. By equating pressures at the level of the oil–water interface, relate the oil column height to the height of water that rises in the right arm. *(Recall that two points at the same height in one connected fluid are at the same pressure, so compare the interface with the point in the right arm at that same level.)*
b. Find the height of that water column.
c. By how much do the two liquid surfaces differ in height, and which is higher?

14.22 ★★★ *[Real-world: recreation.]* A wooden raft of density $600\ \text{kg/m}^3$ measures $3.0\ \text{m} \times 2.0\ \text{m} \times 0.25\ \text{m}$ and floats in fresh

water.

a. Floating empty, what fraction of its thickness is submerged?
b. What is the maximum load (in kg) it can carry before water reaches its top surface?
c. How many $75\ \mathrm{kg}$ people is that?

14.23 ★★★ *[Real-world: engineering.]* A large open tank of water has a small hole of area $1.5\ \mathrm{cm}^2$ in its side, $2.5\ \mathrm{m}$ below the water surface and $0.80\ \mathrm{m}$ above the floor.

a. Using Bernoulli's equation, show that the water leaves the hole at speed $v = \sqrt{2gh}$, and evaluate it. *(Apply equation (14.10) between the top surface and the hole. Both are open to the atmosphere, so the pressures there are equal; and because the tank is wide and the hole small, the surface descends slowly enough to treat its speed as zero.)*
b. Find the volume flow rate through the hole.
c. How far from the base of the tank does the stream strike the floor? Neglect air resistance.

This page titled 14.E: Fluids (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15: ELECTRIC CHARGES AND COULOMB'S LAW

Rub a piece of amber with fur and it will pick up bits of straw. The Greeks knew this twenty-five centuries ago, and their word for amber, *ēlektron*, is where "electricity" and "electron" come from. For most of history the effect was a parlor curiosity, in the same category as the static crackle of a wool sweater, the shock from a door handle on a dry winter day, or a balloon clinging to a wall. It turns out to be the visible edge of one of the two long-range forces in nature, and the one that governs nearly everything we touch.

In Chapters 5 and 6 we met normal forces, friction, tension, and spring forces, and treated them as distinct kinds of push and pull. They are not distinct at all. Every one of them is the *electric* force between the atoms of two surfaces, or between neighboring atoms within a rope or a spring. Apart from gravity, the electric force holds atoms together, binds atoms into molecules, makes solids rigid, and drives every chemical reaction and every nerve impulse. Gravity decides how planets move. Electricity decides almost everything else.

This chapter introduces electric charge, the property of matter that the electric force acts on, and the law that gives the force between two charges. That law, Coulomb's law, has exactly the inverse-square form of Newton's law of gravitation from Chapter 13, so much of what you learned there carries straight over. The one crucial difference is that charge comes in two signs, and that single fact explains why a force that, between fundamental particles, is 10^{36} to 10^{42} times stronger than gravity can stay hidden in everyday life.

Learning Objectives

- Describe the two kinds of electric charge and apply the conservation and quantization of charge.
- Distinguish conductors from insulators, and explain charging by friction, by conduction, and by induction, and why a charged object attracts a neutral one.
- State and apply Coulomb's law for the force between two point charges, in both magnitude and vector form.
- Use the superposition principle to find the net electric force on a charge due to several other charges.
- Describe a continuous charge distribution with a charge density, and integrate to find the force it exerts on a point charge.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

15.1: Electric Charge
15.2: Coulomb's Law
15.3: Discrete Charge Arrangements and the Superposition Principle
15.4: Electric Forces on Continuous Charge Distributions
15.S: Electric Charges and Coulomb's Law (Summary)
15.E: Electric Charges and Coulomb's Law (Problems)

This page titled 15: Electric Charges and Coulomb's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.1: Electric Charge

A few simple experiments reveal the basic facts. Rub two glass rods with silk and hang one from a thread: bring the other near and the hanging rod swings *away*. Rub two plastic rods with fur and they also repel each other. But bring a rubbed glass rod near a rubbed plastic rod and the two *attract*. No experiment ever turns up a third behavior. Every charged object either repels the glass rod and attracts the plastic one, or the reverse.

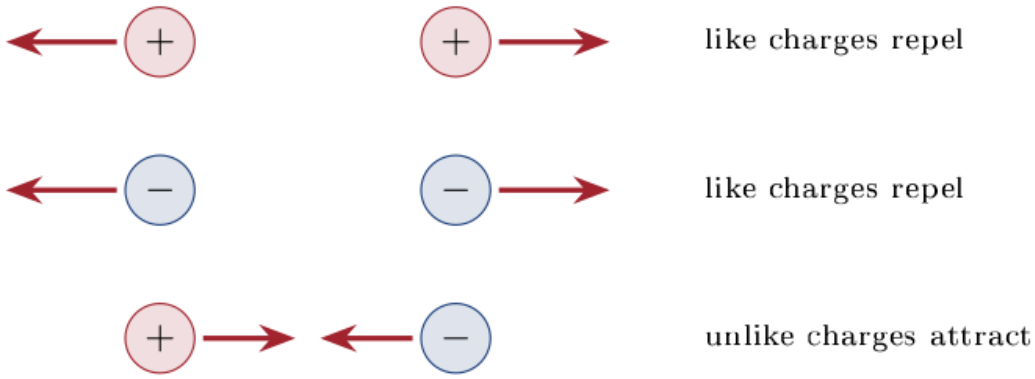


Figure 15.1. The two kinds of charge. Like charges repel and unlike charges attract. In every case the two forces are equal in magnitude and opposite in direction, a Newton's-third-law pair.

•Figure description

Three pairs of charged spheres. Two positive charges push apart, with force arrows pointing outward from each. Two negative charges also push apart. A positive and a negative charge pull together, with force arrows pointing toward each other.

We conclude that **electric charge** is a property of matter that comes in exactly two kinds, and that

like charges repel, and unlike charges attract (Figure 15.1).

Benjamin Franklin named the two kinds **positive** and **negative**, calling the charge on glass rubbed with silk positive. The names are purely conventional, but they were a happy choice, because charges of opposite sign really do behave like positive and negative numbers: equal amounts of each, placed together, add up to an object that exerts no electric force at all. Such an object is **electrically neutral**.

Where charge comes from. Ordinary matter is made of atoms, and each atom is a tiny, dense nucleus of **protons** and **neutrons** surrounded by a cloud of **electrons** (Chapter 41 tells the story of how this picture was discovered). Protons carry positive charge, electrons carry negative charge, and neutrons carry none. The proton's charge and the electron's are *exactly* equal in size; experiments have shown they match to better than one part in 10^{21}. A normal atom has as many electrons as protons and is therefore neutral.

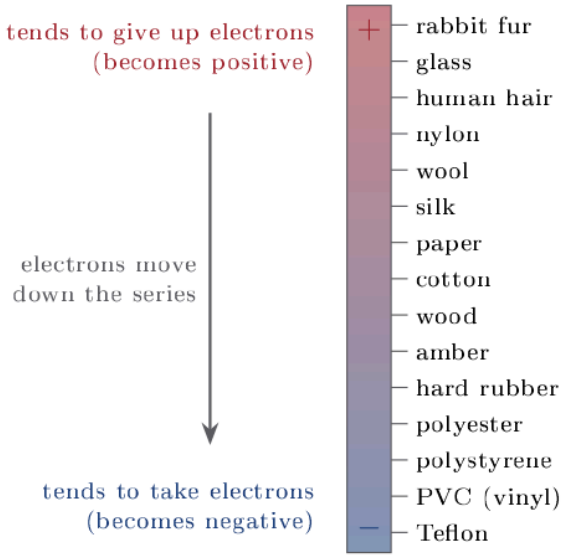


Figure 15.2. A triboelectric series. When two materials are rubbed together, electrons move from the one higher on the list to the one lower down, so the higher material charges positive and the lower one negative. Rubbed glass is positive against silk, and fur leaves a plastic rod negative. The exact order varies somewhat between sources and with surface condition and humidity, so only the general ranking is reliable.

•Figure description

A triboelectric series drawn as a vertical bar shaded from red at the top to blue at the bottom. The top end is labeled: tends to give up electrons, becomes positive. The bottom end is labeled: tends to take electrons, becomes negative. Fifteen materials are listed beside the bar, from top to bottom: rabbit fur, glass, human hair, nylon, wool, silk, paper, cotton, wood, amber, hard rubber, polyester, polystyrene, PVC, and Teflon.

The electrons in the outer part of an atom are held much more loosely than the protons locked in its nucleus, so when two materials are rubbed together, it is electrons that move. Rubbing a balloon on your hair transfers electrons from hair to balloon. The balloon ends up with a surplus of electrons and a net negative charge, and your hair is left with an equal deficit of electrons and a net positive charge. This is **charging by friction**, and which material ends up negative depends on how tightly each holds its electrons. Materials can be ranked in a **triboelectric series** (Figure 15.2): fur, glass, and hair tend to give up electrons, while rubber, plastic, and Teflon tend to take them. When two materials are rubbed together, the one higher in the series gives electrons to the one lower down, so the higher one ends up positive and the lower one negative. The farther apart the two are in the series, the more charge is usually transferred.

The SI unit of charge is the **coulomb** (C). The magnitude of the charge on a proton or an electron is called the **elementary charge**,

$$e = 1.602 \times 10^{-19}\,\mathrm{C}, \tag{15.1.1}$$

so a proton has charge $+e$ and an electron $-e$. Since 2019 the coulomb has been *defined* by fixing e at exactly $1.602176634 \times 10^{-19}$ C. One coulomb is therefore the charge of about 6.24×10^{18} protons. As you will see in the next section, that is an enormous amount of charge. Static electricity in everyday life involves nanocoulombs ($1\ \mathrm{nC} = 10^{-9}\ \mathrm{C}$) to microcoulombs ($1\ \mu\mathrm{C} = 10^{-6}\ \mathrm{C}$).

Charge is conserved. In the balloon-and-hair experiment, no charge was created. Electrons simply moved from one object to the other, and the positive charge left on the hair exactly equals the negative charge gained by the balloon. This is an example of one of the most firmly established laws in physics, the **conservation of charge**: *the net charge of an isolated system never changes.* Charged particles can be created and destroyed. A high-energy photon can turn into an electron and a positron (an anti-electron with charge $+e$), for example. But they are always created and destroyed in combinations whose total charge is zero, so the net charge is unchanged (Chapter 44).

Charge is quantized. Because all ordinary charge is carried by electrons and protons, the net charge on any object is a whole-number multiple of e:

Key Equation: Quantization of charge

$$q = ne, \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{15.1}$$

Charge comes in discrete lumps and cannot be divided more finely than e. This **quantization of charge** was first demonstrated by Robert Millikan in 1909, who watched tiny charged oil droplets float between two charged plates and found that every droplet carried a whole number of elementary charges. You may know that protons and neutrons are made of quarks, which carry charges of $\pm\frac{1}{3}e$ and $\pm\frac{2}{3}e$. Quarks, however, are always bound together in combinations whose total is a whole multiple of e, and no isolated quark has ever been observed (Chapter 44).

In practice, the lumps are so small that we rarely notice them. A charge of $1\ \text{nC}$ is more than six billion elementary charges, and adding or removing one electron changes it by less than a part in a billion. For most purposes we can treat charge as a continuous quantity, just as we treat water as a smooth fluid even though it is made of molecules. We will use that idea in Section 15.4.

Example 15.1: Charging a balloon

After being rubbed on someone's hair, a rubber balloon carries a charge of $-45.0\ \text{nC}$. (a) How many excess electrons are on the balloon? (b) By how much has the balloon's mass increased? (The mass of an electron is $m_e = 9.11 \times 10^{-31}\ \text{kg}$.)

Solution

(a) From equation (15.1), the number of electrons is the size of the charge divided by the charge per electron:

$$n = \frac{|q|}{e} = \frac{45.0 \times 10^{-9}\,\text{C}}{1.602 \times 10^{-19}\,\text{C}} = \boxed{2.81 \times 10^{11}}. \tag{15.1.2}$$

(b) Each electron brings its mass along:

$$\Delta m = nm_e = (2.81 \times 10^{11})(9.11 \times 10^{-31}\,\text{kg}) \approx \boxed{2.6 \times 10^{-19}\,\text{kg}}. \tag{15.1.3}$$

Some 280 billion electrons sounds like a lot, but they weigh far less than a single bacterium. The electrical effects of charging are dramatic, while the change in mass is utterly negligible, and we will ignore it from now on. Note, too, how small a fraction of the balloon's electrons have moved: a few grams of rubber contain roughly 10^{24} electrons, so the balloon has gained only about one extra electron for every few trillion it already had.

Example 15.2: The hidden charge in a bottle of water

How much negative charge is carried by all the electrons in $1.00\ \text{L}$ of water? Each water molecule, H_2O, has 10 electrons (1 from each hydrogen and 8 from the oxygen), and the molar mass of water is $18.02\ \text{g/mol}$.

Solution

The $1000\ \text{g}$ of water contains $1000/18.02 = 55.5\ \text{mol}$ of molecules, or

$$N = (55.5\,\text{mol})(6.022 \times 10^{23}\,\text{mol}^{-1}) = 3.34 \times 10^{25}\ \text{molecules}. \tag{15.1.4}$$

With 10 electrons per molecule, the total electron charge is

$$q = -10Ne = -(10)(3.34 \times 10^{25})(1.602 \times 10^{-19}\,\text{C}) \approx \boxed{-5.4 \times 10^{7}\,\text{C}}. \tag{15.1.5}$$

Fifty million coulombs. The water is neutral, of course, because its protons carry exactly $+5.4 \times 10^{7}\ \text{C}$. But the number shows how much charge ordinary matter holds in perfect balance. Charging the balloon of Example 15.1 upset the same kind of balance in its few grams of rubber by less than one part in a trillion.

Note: Why "negative" electrons?

Franklin chose the signs in the 1740s, a century and a half before anyone knew about electrons. His choice happened to make the particle that actually moves in most situations the negative one. Nothing about the physics would change if the labels were swapped, but the choice has one lasting consequence. The direction of electric current was defined as the direction positive charge flows, and in a metal wire the electrons move the *opposite* way. We will return to this in Chapter 20.

Checkpoint 15.1

A glass rod rubbed with silk acquires a charge of $+12\ \mathrm{nC}$. (a) What is the charge on the silk, and how do you know? (b) Did the rod gain or lose electrons, and how many? (c) A student claims that a tiny dust grain carries a charge of $4.0 \times 10^{-19}\ \mathrm{C}$. Without any further information, explain why this measurement must be wrong.

Answer

Conductors and Insulators

Rub a plastic rod and the charge stays where you put it: one end can be charged while the other end stays neutral. Try the same with a copper rod held in your hand and it seems not to charge at all. The difference lies in how freely electrons can move through each material.

In a **conductor**, such as a metal, some of the electrons are not bound to any particular atom. Typically one or two electrons per atom are free to wander throughout the whole piece of metal. These **free electrons** (also called conduction electrons) move in response to even a tiny electric force, so charge placed on a conductor spreads out over it almost instantly. When you rub a copper rod held in your hand, the charge you deposit flows through the rod, through your body, and into the ground. You yourself are a fair conductor, because your body is mostly salt water. Aqueous solutions of salts, acids, and bases conduct, as does the ionized gas called a plasma.

In an **insulator**, such as glass, rubber, plastic, or dry wood, every electron is tightly bound to its own atom or molecule. Charge placed on an insulator stays put. This is why static electricity is easiest to see with insulators, and why the handles of electrical tools are made of plastic.

Between the two lie the **semiconductors**, such as silicon and germanium. A pure semiconductor has very few free electrons, but adding tiny, carefully controlled amounts of impurities can change its conductivity by many orders of magnitude. That controllability is what makes transistors, and with them all of modern electronics, possible.

The Earth itself is a large conductor with an effectively unlimited supply of electrons to give or take. Connecting a conductor to the Earth with a wire is called **grounding** it. A grounded conductor can gain or lose as many electrons as needed to reach a neutral (or nearly neutral) state.

Charging by conduction and by induction. There are two ways to put charge on a conductor using an already-charged object. The obvious one is to touch them together. If a negatively charged rod touches a neutral metal sphere, some of the rod's excess electrons move onto the sphere and spread over its surface, and the sphere ends up negative, with the same sign as the rod. This is **charging by conduction**, or by contact.

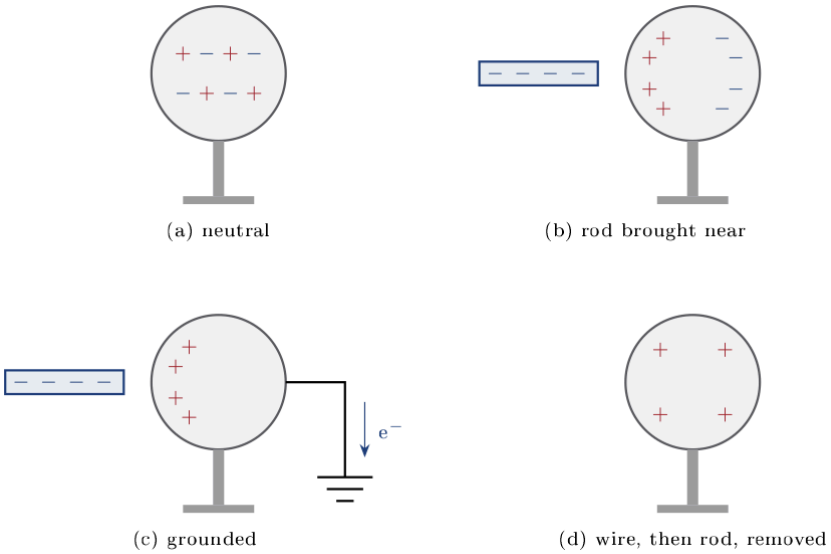


Figure 15.3. Charging a conductor by induction. (a) A neutral metal sphere. (b) A negative rod repels the sphere's free electrons to the far side. (c) Grounding the far side lets those electrons escape to the Earth. (d) With the ground wire removed first and the rod second, the sphere is left with a net *positive* charge, opposite in sign to the rod, which never touched it.

•Figure description

Four panels showing a metal sphere on an insulating stand being charged by induction. In panel a the neutral sphere has positive and negative charges mixed evenly. In panel b a negatively charged rod is held to the left of the sphere; the sphere's negative charges have moved to the right side, leaving positive charges on the left side. In panel c a wire connects the right side of the sphere to ground, and electrons flow down the wire into the ground. In panel d the wire and the rod have been removed and the sphere carries positive charge spread evenly around its surface.

The second way works without any contact at all. Figure 15.3 shows the steps:

a. A metal sphere sits on an insulating stand. It is neutral, with equal numbers of positive and negative charges.
b. A negatively charged rod is brought near, without touching. It repels the sphere's free electrons, which crowd to the far side, leaving the near side with a deficit of electrons and so a net positive charge. The sphere as a whole is still neutral. Its charge has only been rearranged.
c. With the rod still in place, the far side is connected to ground by a wire. The repelled electrons now have somewhere to go, and they flow down the wire into the Earth.
d. The ground wire is disconnected, and *then* the rod is removed. The electrons that left cannot return, so the sphere has a net positive charge, which spreads evenly over its surface.

This is **charging by induction**. Notice that the sphere ends up with the *opposite* sign to the rod, and that the rod loses none of its own charge in the process, so it can be used again and again. The order of the last step matters: if the rod is taken away while the ground wire is still attached, electrons flow back up from the ground and the sphere returns to neutral.

Polarization, and why neutral objects are attracted. Step (b) of Figure 15.3 already shows something surprising. Before any grounding, the neutral sphere is *attracted* to the rod. Its near side is positive and its far side negative, so the rod pulls on the near side and pushes on the far side. The near side is closer, the electric force weakens with distance (as the next section shows), and so the pull wins. A separation of positive and negative charge within an object is called **polarization**, and a polarized neutral object is always drawn toward a nearby charge, *whatever that charge's sign*. A positive rod would pull the electrons to the near side instead, and the result would again be attraction.

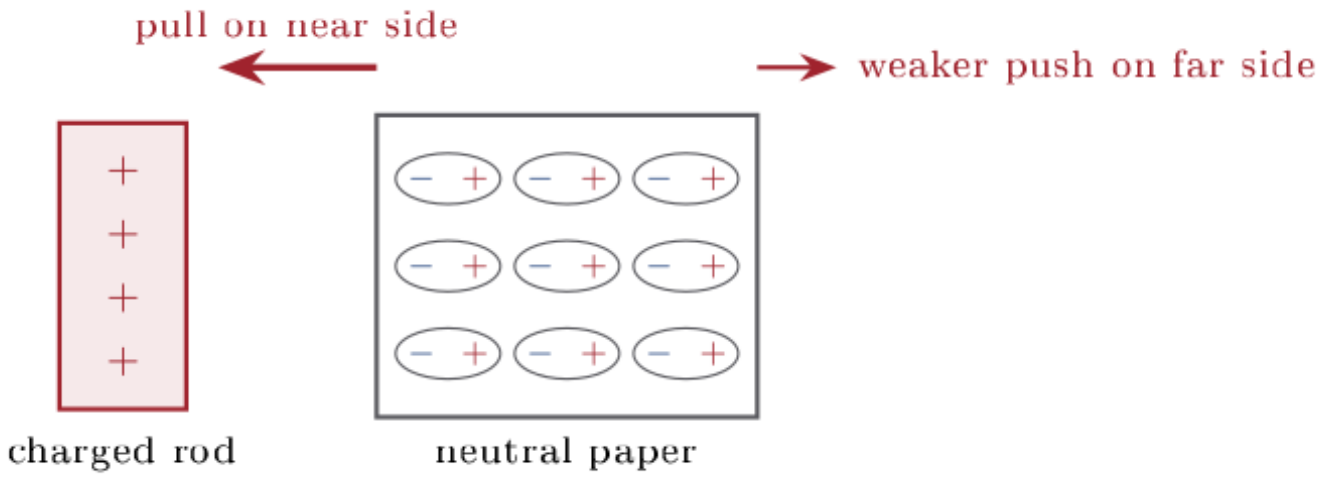


Figure 15.4. Polarization of an insulator. A nearby charge slightly separates the charges within each molecule, leaving a layer of opposite charge on the near surface and like charge on the far surface. Because the near layer is closer, its attraction outweighs the repulsion of the far layer, and the neutral object is drawn toward the rod.

•Figure description

A positively charged rod on the left is held near a neutral scrap of paper on the right. Inside the paper, each molecule is drawn as a small oval with its negative end toward the rod and its positive end away from it. A thin layer of negative charge therefore faces the rod and a layer of positive charge lies on the far side. A large force arrow on the near side points toward the rod and a smaller arrow on the far side points away, so the net force pulls the paper toward the rod.

Insulators polarize too, even though their electrons cannot travel. Within each atom or molecule, the electron cloud shifts slightly toward a nearby positive charge, or away from a negative one, so that every molecule becomes a tiny separated pair of charges (Figure 15.4). Many molecules, water among them, are permanently polarized to begin with, and a nearby charge simply swings them into line. Either way, the surface nearest the charge ends up with a thin layer of opposite charge. This is how a charged comb picks up scraps of paper, and why a rubbed balloon clings to a neutral wall.

Checkpoint 15.2

You have two identical metal spheres on insulating stands, touching each other, and a positively charged rod. (a) Describe how you could use the rod to give the two spheres charges that are equal in size and opposite in sign, without the rod ever touching either one. Which sphere ends up with which sign? (b) Suppose instead that you let the rod touch one of the spheres, then separate the spheres. What charges do they end up with?

Answer

Answers to Checkpoints

Checkpoint 15.1. (a) $-12\ \mathrm{nC}$. Charge is conserved: rubbing only moves charge from one object to the other, so whatever the rod gains, the silk loses. (b) A positive charge means the rod *lost* electrons, $N = Q/e = (12 \times 10^{-9}\ \mathrm{C})/(1.602 \times 10^{-19}\ \mathrm{C}) \approx 7.5 \times 10^{10}$ of them. (c) Charge is quantized: every charge is a whole-number multiple of $e = 1.602 \times 10^{-19}\ \mathrm{C}$. But $4.0 \times 10^{-19}/1.602 \times 10^{-19} \approx 2.5$, not a whole number, so no real object can carry that charge.

Checkpoint 15.2. (a) Bring the rod near sphere A, the sphere on the rod's side, without touching. Its positive charge attracts electrons, which flow through the touching spheres toward it: sphere A becomes negative and sphere B, the far one, is left positive. Keeping the rod in place, pull the spheres apart, then take the rod away. A keeps a negative charge and B an equal positive charge, since the total was zero throughout. (b) Touching lets some electrons flow from the spheres onto the rod, so the pair is left with a net *positive* charge. Because the spheres are identical and were in contact, that charge divides equally: both end up positive, with equal charges.

This page titled 15.1: Electric Charge is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.2: Coulomb's Law

In 1785 the French engineer Charles-Augustin de Coulomb measured the force between two small charged spheres with a torsion balance, the same kind of instrument Henry Cavendish would use thirteen years later to measure the gravitational constant (Section 13.1). Coulomb hung a light insulating rod from a thin fiber, with a small charged sphere at one end, and brought a second charged sphere near it. From the angle through which the fiber twisted, he could read off the force. He found that the force is inversely proportional to the square of the distance between the spheres, and proportional to the product of their charges.

For two **point charges**, meaning charged objects whose sizes are small compared with the distance between them, the result is **Coulomb's law**:

Key Equation: Coulomb's law

$$F_e = k_e \frac{|q_1|\,|q_2|}{r^2}. \tag{15.2}$$

Here F_e is the magnitude of the force that each charge exerts on the other, and r is the distance between them. The force is directed along the line joining the two charges. It is repulsive if the charges have the same sign and attractive if they have opposite signs. Like gravity, it obeys Newton's third law: the force on q_1 due to q_2 is equal in magnitude and opposite in direction to the force on q_2 due to q_1, even if one charge is much larger than the other.

The proportionality constant k_e is called the **Coulomb constant**. In SI units it is

$$k_e = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9\,\mathrm{N\cdot m^2/C^2}, \tag{15.2.1}$$

where

$$\epsilon_0 = 8.85 \times 10^{-12}\,\mathrm{C^2/(N\cdot m^2)} \tag{15.2.2}$$

is the **permittivity of free space**. Writing the constant as $1/4\pi\epsilon_0$ looks like an unnecessary complication for now. It pays off in Chapter 17, where the 4π (the solid angle of a whole sphere) cancels and leaves Gauss's law in a very simple form. You will meet ϵ_0 again in Chapter 27 (no spoilers yet).

Note: How big is a coulomb?

The size of k_e tells you that a coulomb is a lot of charge. Two point charges of $1\ \mathrm{C}$ each, placed $1\ \mathrm{m}$ apart, would push on each other with a force of $8.99 \times 10^9\ \mathrm{N}$, about the weight of a million tonnes. Compare the gravitational constant, $G = 6.674 \times 10^{-11}\ \mathrm{N\cdot m^2/kg^2}$: two $1\ \mathrm{kg}$ masses a meter apart attract with less than a ten-billionth of a newton. This is why the charges in electrostatics problems are usually measured in microcoulombs or nanocoulombs, and why moving even a trillionth of an object's electrons produces effects you can see.

Coulomb's law and gravity. Coulomb's law has exactly the same form as Newton's law of gravitation, equation (13.1), with charges taking the place of masses. Everything Chapter 13 said about the inverse-square law applies here too: the force falls to a quarter when the distance doubles, it never quite reaches zero, and r is measured center to center. There is also an electrical version of the shell theorem. A *spherically symmetric* distribution of charge acts on any charge outside it as though all of its charge were concentrated at the center. Chapter 17 proves this. Until then, we will simply use it to treat small charged spheres as point charges.

The differences matter just as much. Gravity is always attractive, but the electric force can attract or repel. And the electric force is far, far stronger, as the next example shows. These two facts together explain why the stronger force is the less obvious one. Because every large body contains almost exactly equal amounts of positive and negative charge, electric forces between large bodies nearly cancel. Gravity has no negative mass to cancel it, so it simply adds up, and over astronomical distances it wins.

Example 15.3: Electricity versus gravity in a hydrogen atom

In a hydrogen atom, the electron and the proton are, on average, about $r = 5.29 \times 10^{-11}\ \text{m}$ apart. Compare the electric and gravitational forces between them. (Use $m_e = 9.11 \times 10^{-31}\ \text{kg}$ and $m_p = 1.673 \times 10^{-27}\ \text{kg}$.)

Solution

Both particles carry charge of magnitude e, so by equation (15.2)

$$F_e = k_e \frac{e^2}{r^2} = \frac{(8.99 \times 10^9)(1.602 \times 10^{-19})^2}{(5.29 \times 10^{-11})^2} = 8.24 \times 10^{-8}\,\text{N}. \tag{15.2.3}$$

The gravitational force, from equation (13.1), is

$$F_g = G\frac{m_e m_p}{r^2} = \frac{(6.674 \times 10^{-11})(9.11 \times 10^{-31})(1.673 \times 10^{-27})}{(5.29 \times 10^{-11})^2} = 3.63 \times 10^{-47}\,\text{N}. \tag{15.2.4}$$

Both forces are attractive. Their ratio is

$$\frac{F_e}{F_g} = \frac{8.24 \times 10^{-8}}{3.63 \times 10^{-47}} \approx \boxed{2.3 \times 10^{39}}. \tag{15.2.5}$$

Because both forces are inverse-square, r cancels from the ratio. It is the same at every separation. It does depend on which particles you compare, because the masses differ while the charges do not: two protons give about 1.2×10^{36}, and two electrons, the lightest charged pair, about 4×10^{42}. Between fundamental particles, then, the electric force is 10^{36} to 10^{42} times stronger than gravity. Within atoms and molecules gravity is utterly irrelevant, and the structure of all matter is set by the electric force alone.

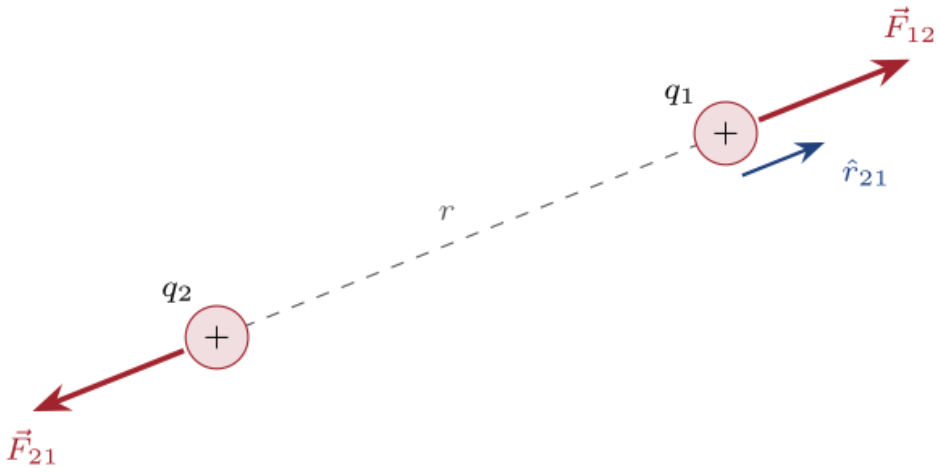


Figure 15.5. Coulomb's law in vector form. The unit vector $\hat{r}_{21}$ points from q_2 toward q_1. For like charges, as drawn, $\mathbf{F}_{12}$ points along $\hat{r}_{21}$ and the charges repel. For unlike charges $q_1 q_2 < 0$, and both forces reverse, pointing toward the other charge.

•Figure description

Two positive point charges, q sub 2 at the lower left and q sub 1 at the upper right, a distance r apart along a dashed line. A short unit vector r hat sub 2 1 starts at q sub 1 and points directly away from q sub 2. The force F sub 1 2 on q sub 1 points along that unit vector, away from q sub 2, and the force F sub 2 1 on q sub 2 points in the opposite direction, away from q sub 1.

Coulomb's law as a vector. Equation (15.2) gives only the magnitude, and you supply the direction by asking whether the charges attract or repel. It is sometimes convenient to let the signs of the charges do that work. We use the same subscript convention as Chapter 13: $\mathbf{F}_{12}$ is the force *on* charge 1 *due to* charge 2. Let $\hat{r}_{21}$ be the unit vector that points *from* q_2 *toward* q_1 (Figure 15.5). Then

Key Equation: Coulomb's law, vector form

$$\mathbf{F}_{12} = k_e \frac{q_1 q_2}{r^2} \hat{r}_{21}. \tag{15.3}$$

If q_1 and q_2 have the same sign, the product $q_1 q_2$ is positive and $\mathbf{F}_{12}$ points along $\hat{r}_{21}$, away from q_2: repulsion. If they have opposite signs, $q_1 q_2$ is negative and $\mathbf{F}_{12}$ points back toward q_2: attraction. The force on q_2 is $\mathbf{F}_{21} = -\mathbf{F}_{12}$, as Newton's third law requires.

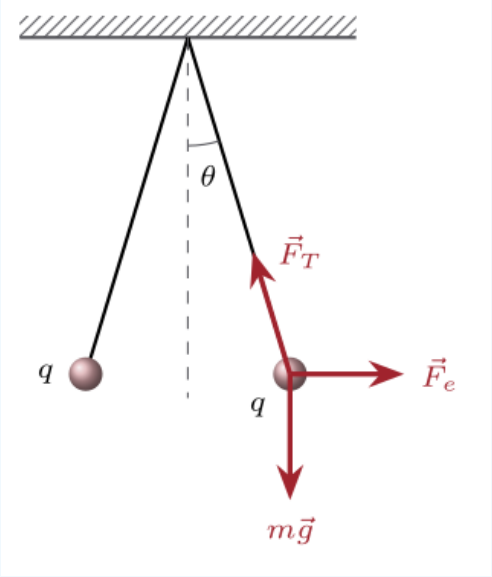
Note: Magnitudes first, directions by inspection

The vector form is compact and is the natural starting point for the integrals of Section 15.4. For problems worked by hand, though, most physicists use equation (15.2) with the *magnitudes* of the charges, and then decide the direction by asking one question: do these two charges attract or repel? Draw the force on your sketch, pointing toward or away from the other charge, and only then break it into components. Mixing the two approaches, by putting signed charges into the magnitude formula, is the most common source of sign errors in this chapter.

Coulomb's law as stated here is a law of *electrostatics*: it gives the force between charges that are at rest, or moving slowly. Charges in motion also exert *magnetic* forces on one another, which are the subject of Chapter 22.

Example 15.4: Charged spheres hanging from threads

Two identical small spheres, each of mass $m = 2.00\ \text{g}$, hang from light insulating threads of length $L = 0.500\ \text{m}$ attached to the same point. When the spheres are given equal charges q, they repel and come to rest with each thread making an angle $\theta = 10.0^\circ$ with the vertical. Find q.



•Figure description
Two small charged spheres hang from threads attached to a common point on the ceiling. The threads spread apart symmetrically, each at angle theta from a dashed vertical line. On the right-hand sphere three forces are drawn: the tension F sub T along the thread toward the attachment point, the weight m g straight down, and the electric force F sub e horizontally to the right, away from the other sphere.

Solution

Each sphere is in equilibrium under three forces: the tension $\mathbf{F}_T$ along its thread, its weight $m\mathbf{g}$, and the electric force $\mathbf{F}_e$ from the other sphere, which is horizontal. For the right-hand sphere, Newton's second law with $\mathbf{a} = 0$ gives

$$\sum F_x : \ F_e - F_T \sin\theta = 0, \qquad \sum F_y : \ F_T \cos\theta - mg = 0. \tag{15.2.6}$$

Dividing the first equation by the second eliminates F_T:

$$F_e = mg\tan\theta = (2.00\times10^{-3}\,\text{kg})(9.80\,\text{m/s}^2)\tan 10.0^\circ = 3.46\times10^{-3}\,\text{N}. \tag{15.2.7}$$

The distance between the spheres is $r = 2L\sin\theta = 2(0.500\ \text{m})\sin 10.0^\circ = 0.174\ \text{m}$. Solving Coulomb's law, $F_e = k_e q^2/r^2$, for q:

$$q = r\sqrt{\frac{F_e}{k_e}} = (0.174\,\text{m})\sqrt{\frac{3.46\times10^{-3}\,\text{N}}{8.99\times10^9\,\text{N}\cdot\text{m}^2/\text{C}^2}} \approx \boxed{1.08\times10^{-7}\,\text{C}} = 108\,\text{nC}. \tag{15.2.8}$$

We can find only the *size* of q. The spheres would hang exactly the same way if both were negative. A charge of about $0.1\ \mu\text{C}$ is typical of what you can produce by rubbing.

Checkpoint 15.3

Two point charges exert a force of $0.36\ \text{N}$ on each other. What is the force if (a) the distance between them is doubled; (b) both charges are doubled and the distance is tripled; (c) one charge is replaced by a charge of the same size but opposite sign? In each case, say whether the force is attractive or repulsive compared with the original.

Answer

Answers to Checkpoints

Checkpoint 15.3. Coulomb's law gives $F \propto |q_1 q_2|/r^2$. (a) Doubling r divides the force by $2^2 = 4$: $0.090\ \text{N}$, of the same type as before. (b) Doubling both charges multiplies by $2\times2 = 4$; tripling r divides by 9: $F = (0.36)(4/9) = 0.16\ \text{N}$, again of the same type. (c) The size is unchanged, $0.36\ \text{N}$, but the type reverses: if the charges repelled before, they now attract, and vice versa.

This page titled 15.2: Coulomb's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.3: Discrete Charge Arrangements and the Superposition Principle

When several charges act on one charge, what is the net force? Experiment gives a simple answer. The force that any one charge exerts on another is exactly what Coulomb's law predicts for that pair, whether or not other charges are present. The net force is then the *vector sum* of the individual forces. This is the **superposition principle** for electric forces. For a charge q_1 acted on by charges $q_2, q_3, \dots, q_N$,

Key Equation: Superposition of electric forces

$$\mathbf{F}_1 = \mathbf{F}_{12} + \mathbf{F}_{13} + \cdots + \mathbf{F}_{1N} = \sum_{j=2}^{N} \mathbf{F}_{1j}. \tag{15.4}$$

You have already used this idea for gravitational forces in Chapter 13, and for forces of every kind in Chapter 5. It is worth pausing to notice that it is a genuine experimental fact, not a mathematical necessity. It holds because the electric force between two charges does not depend on anything else around them, and it holds to extraordinary precision for everything in this book. Note that a charge never exerts a force on itself, so the sum leaves out $j = 1$.

Note: A strategy for superposition problems

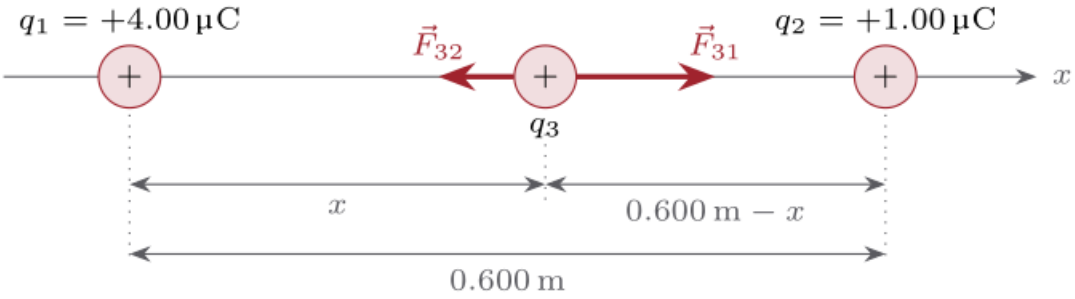
1. Draw a sketch showing every charge, and choose a coordinate system. Put the charge you are finding the force on at a convenient place, often the origin.
2. For each of the other charges, find the *magnitude* of its force from equation (15.2), using the magnitudes of the charges and the distance to that charge.
3. Decide each force's direction (toward the other charge if they attract, away from it if they repel) and draw it on your sketch.
4. Resolve each force into x and y components, exactly as in Chapter 3, and add the components.
5. Look for symmetry before you calculate. Equal charges at equal distances often produce components that cancel, and spotting that saves a lot of work.

Example 15.5: Where is the net force zero?

A charge $q_1 = +4.00\ \mu\text{C}$ is fixed at the origin and a charge $q_2 = +1.00\ \mu\text{C}$ is fixed at $x = 0.600\ \text{m}$. Where on the x-axis can a third charge q_3 be placed so that the net electric force on it is zero?

Solution

First decide which region can work. To the left of q_1 or to the right of q_2, both charges push (or both pull) q_3 in the same direction, so the forces cannot cancel. Between them, the two forces on q_3 point in opposite directions, whatever the sign of q_3. Let q_3 sit at position x between 0 and $0.600\ \text{m}$, as in the sketch, where q_3 is drawn positive. The forces balance when their magnitudes are equal:



•Figure description
A horizontal x-axis. A positive charge q sub 1 of 4.00 microcoulombs sits at the origin, and a positive charge q sub 2 of 1.00 microcoulomb sits 0.600 meters to its right. A third charge q sub 3, drawn as positive, sits on the axis between them, a distance x from q sub 1 and a distance 0.600 meters minus x from q sub 2. Two forces act on q sub 3: F sub 3 1, from q sub 1, points to the right, away from q sub 1, and F sub 3 2, from q sub 2, points to the left, away from q sub 2. The two forces point in opposite directions.

$$k_e \frac{|q_1||q_3|}{x^2} = k_e \frac{|q_2||q_3|}{(0.600 - x)^2}. \tag{15.3.1}$$

Both k_e and $|q_3|$ cancel. Taking the square root of both sides,

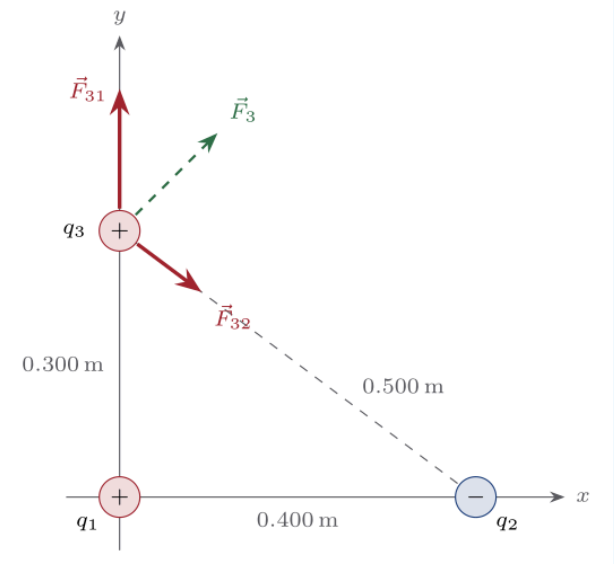
$$\frac{\sqrt{4.00}}{x} = \frac{\sqrt{1.00}}{0.600 - x} \implies 2(0.600 - x) = x \implies x = \boxed{0.400\,\text{m}}. \tag{15.3.2}$$

The balance point is twice as far from q_1 as from q_2, because q_1 is four times larger and the force falls off as the square of distance. The answer does not depend on the size or sign of q_3.

The balance point in this example is an equilibrium, but it is a delicate one. If q_3 is positive and is nudged along the axis toward q_2, the repulsion from q_2 grows and pushes it back, so along the axis the equilibrium is stable. But nudge it slightly *off* the axis and both charges push it farther away: sideways, the equilibrium is unstable. (For a negative q_3 the situation reverses, and it is unstable along the axis.) This is not a coincidence of this example. **Earnshaw's theorem** states that no arrangement of fixed charges can hold another charge in stable equilibrium by electric forces alone. Every equilibrium is unstable in at least one direction. This is one reason the classical picture of electrons sitting at rest inside atoms could never have worked.

Example 15.6: Net force on a charge in two dimensions

Three point charges lie at the corners of a right triangle, as shown: $q_1 = +2.00\ \mu\text{C}$ at the origin, $q_2 = -4.00\ \mu\text{C}$ at $(0.400\ \text{m},\ 0)$, and $q_3 = +5.00\ \mu\text{C}$ at $(0,\ 0.300\ \text{m})$. Find the net electric force on q_3.



•Figure description

A right triangle of three charges. q sub 1, positive, sits at the origin; q sub 2, negative, sits 0.400 meters to its right on the x-axis; q sub 3, positive, sits 0.300 meters above q sub 1 on the y-axis. The side from q sub 3 to q sub 2 is 0.500 meters long. Two forces act on q sub 3: F sub 3 1 points straight up, away from q sub 1, and a shorter force F sub 3 2 points down and to the right, toward q sub 2. Their dashed resultant F sub 3 points up and to the right at about 45 degrees.

Solution

Find the magnitude and direction of each force on q_3 separately.

Force from q_1. The distance is $0.300\ \text{m}$, and the two positive charges repel, so $\mathbf{F}_{31}$ points straight up, in the $+y$ direction:

$$F_{31} = k_e \frac{|q_3||q_1|}{r_{31}^2} = \frac{(8.99\times10^9)(5.00\times10^{-6})(2.00\times10^{-6})}{(0.300)^2} = 0.999\,\text{N}. \tag{15.3.3}$$

Force from q_2. The distance is $\sqrt{0.300^2+0.400^2} = 0.500\ \text{m}$. The charges have opposite signs and attract, so $\mathbf{F}_{32}$ points from q_3 toward q_2, along the unit vector $(0.400\ \hat{\imath} - 0.300\ \hat{\jmath})/0.500 = 0.800\ \hat{\imath} - 0.600\ \hat{\jmath}$:

$$F_{32} = k_e \frac{|q_3||q_2|}{r_{32}^2} = \frac{(8.99\times10^9)(5.00\times10^{-6})(4.00\times10^{-6})}{(0.500)^2} = 0.719\,\text{N}. \tag{15.3.4}$$

Add components.

$$\begin{aligned} F_{3x} &= 0 + (0.800)(0.719\,\text{N}) = 0.575\,\text{N}, \\ F_{3y} &= 0.999\,\text{N} - (0.600)(0.719\,\text{N}) = 0.567\,\text{N}. \end{aligned} \tag{15.3.5}$$

The net force has magnitude and direction

$$F_3 = \sqrt{F_{3x}^2 + F_{3y}^2} = \boxed{0.808\,\text{N}}, \qquad \theta = \arctan\frac{F_{3y}}{F_{3x}} = \boxed{44.6^\circ}\ \text{above the}\ +x\text{-axis}. \tag{15.3.6}$$

Although q_2 carries twice the charge of q_1, its force is smaller, because it is $5/3$ as far away and $(5/3)^2 \approx 2.8$.

Checkpoint 15.4

Four equal positive charges sit at the corners of a square, and a fifth positive charge is placed at the exact center. (a) Without calculating, what is the net force on the center charge, and why? (b) One of the corner charges is removed. Now which way does the net force on the center charge point? *(Hint: removing a charge is the same as adding an equal and opposite charge at that corner.)*

Answer

Answers to Checkpoints

Checkpoint 15.4. (a) Zero. The four corner charges are equal and equidistant from the center, and they sit in opposite pairs, so each push is canceled by an equal push from the opposite corner. (b) Removing one corner charge is equivalent to leaving all four in place (net force zero) and adding an equal *negative* charge at that corner. That negative charge attracts the positive center charge, so the net force points diagonally toward the empty corner. Equivalently, the charge at the opposite corner is no longer balanced and pushes the center charge toward the vacancy.

This page titled 15.3: Discrete Charge Arrangements and the Superposition Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.4: Electric Forces on Continuous Charge Distributions

A charged rod, plate, or sphere contains so many elementary charges that summing Coulomb's law over each one is out of the question. We have seen that at this scale the graininess of charge is invisible, so instead we treat the charge as spread smoothly through the object, a **continuous charge distribution**. The sum of equation (15.4) then becomes an integral, just as the sum over particles that gave the moment of inertia became an integral in Chapter 10.

Charge density. To describe how charge is spread through an object, we use a **charge density**, the amount of charge per unit length, area, or volume, depending on the shape:

- For charge along a thin rod or wire, the **linear charge density** λ is the charge per unit length, in $\mathrm{C/m}$. A small piece of length $d\ell$ carries charge $dq = \lambda\ d\ell$.
- For charge spread over a thin sheet or the surface of a conductor, the **surface charge density** σ is the charge per unit area, in $\mathrm{C/m^2}$. A small patch of area dA carries $dq = \sigma\ dA$.
- For charge spread through a volume, the **volume charge density** ρ is the charge per unit volume, in $\mathrm{C/m^3}$. A small element of volume dV carries $dq = \rho\ dV$.

If the charge Q is spread *uniformly*, the densities are simply

Key Equation: Uniform charge densities

$$\lambda = \frac{Q}{L}, \qquad \sigma = \frac{Q}{A}, \qquad \rho = \frac{Q}{V}. \tag{15.5}$$

(The symbol ρ does double duty as mass density, as in Chapter 14. Context always makes clear which one is meant.)

The force from a continuous distribution. Divide the distribution into tiny elements, each carrying charge dq and small enough to be treated as a point charge. By Coulomb's law, the element exerts a force $d\mathbf{F}$ on a point charge q a distance r away. By superposition, the total force is the sum, in the limit of infinitely many tiny elements, of all these contributions:

Key Equation: Force from a continuous charge distribution

$$\mathbf{F} = \int d\mathbf{F} = k_e\, q \int \frac{dq}{r^2}\,\hat{r}. \tag{15.6}$$

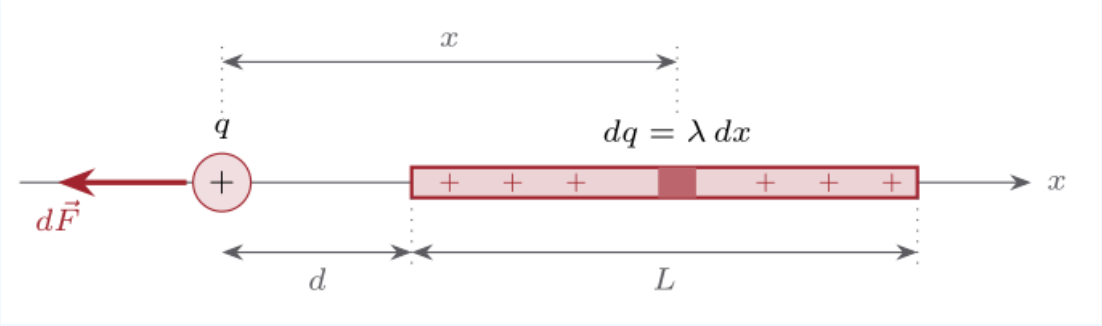
Here $\hat{r}$ is the unit vector pointing from the element dq toward the point charge q, as in equation (15.3), and both r and $\hat{r}$ generally change from one element to the next. Since $d\mathbf{F}$ is a vector, equation (15.6) is really two or three integrals, one for each component. In practice the work always follows the same steps:

1. Sketch the distribution and the point charge. Pick a typical element dq, somewhere in the middle of the distribution rather than at a special point.
2. Write dq in terms of the charge density and a coordinate: $dq = \lambda\ dx$, for example.
3. Write the distance r from dq to q, and the direction of $d\mathbf{F}$, in terms of that coordinate.
4. Use symmetry to decide which components cancel when all the elements are added, and integrate only the ones that survive.
5. Check the answer in a limit you already understand, usually far away, where the distribution should look like a point charge.

The next two examples carry out these steps for the two simplest cases. In the first, every element pushes in the same direction, so only magnitudes need adding. In the second, symmetry does most of the work.

Example 15.7: A point charge on the axis of a charged rod

A thin rod of length L carries a total charge Q spread uniformly along it. A point charge q sits on the rod's axis, a distance d beyond one end, as shown. (a) Find the electric force on q. (b) Evaluate it for $L = 0.200\ \mathrm{m}$, $d = 0.100\ \mathrm{m}$, $Q = +3.00\ \mu\mathrm{C}$, and $q = +1.00\ \mu\mathrm{C}$.



•Figure description
A point charge q sits at the origin on the left. A uniformly charged rod lies along the x-axis to its right, beginning a distance d from q and extending for a length L. A small element of the rod of width d x, at distance x from q, is highlighted and labeled d q equals lambda d x. The force d F that this element exerts on q is drawn as an arrow pointing to the left, directly away from the rod.

Solution

(a) Put q at the origin and let the rod lie along the x-axis from $x = d$ to $x = d + L$. The linear charge density is $\lambda = Q/L$, so an element of length dx at position x carries charge $dq = \lambda\ dx$ and is a distance x from q. Taking Q and q to have the same sign, every element pushes q in the $-x$ direction, away from the rod, so the contributions all point the same way and we can simply add their magnitudes:

$$dF = k_e \frac{q\,dq}{x^2} = k_e\, q\lambda\, \frac{dx}{x^2}. \tag{15.4.1}$$

Integrating over the length of the rod,

$$F = k_e\, q\lambda \int_d^{d+L} \frac{dx}{x^2} = k_e\, q\lambda \left[-\frac{1}{x}\right]_d^{d+L} = k_e\, q\lambda \left(\frac{1}{d} - \frac{1}{d+L}\right) = k_e\, q\lambda\, \frac{L}{d(d+L)}. \tag{15.4.2}$$

With $\lambda L = Q$, this is

$$F = \frac{k_e\, qQ}{d(d+L)}, \tag{15.7}$$

directed along the rod's axis, away from the rod for like charges and toward it for unlike charges.

Check. Far from the rod, $d \gg L$, so $d + L \approx d$ and $F \to k_e qQ/d^2$: from far away the rod looks like a point charge, as it should.

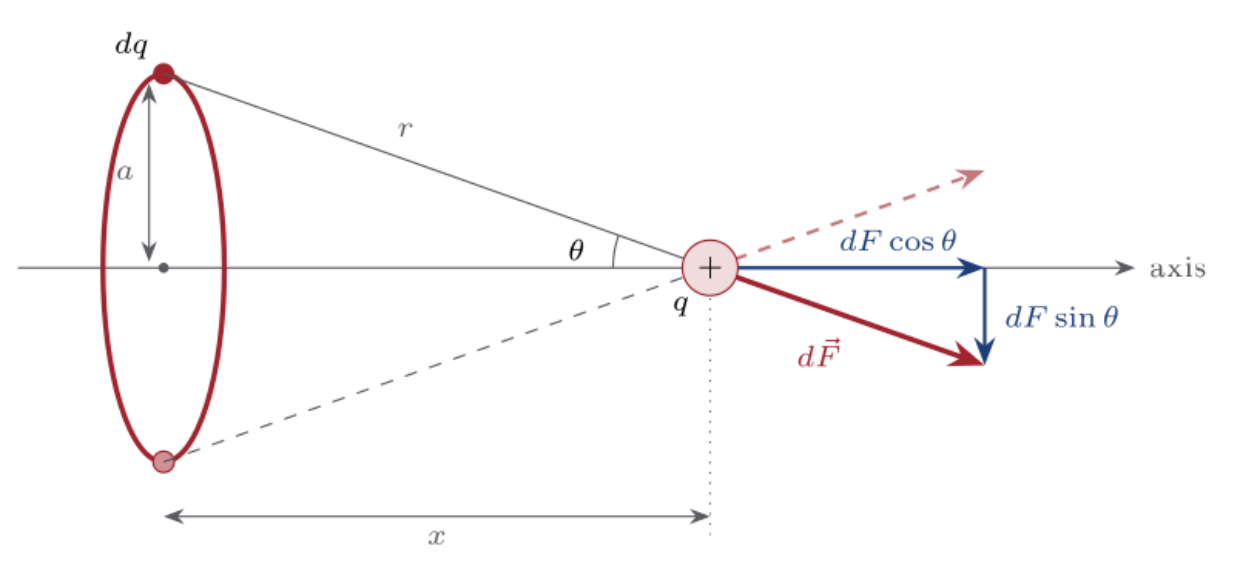
(b) Substituting the numbers,

$$F = \frac{(8.99 \times 10^9)(1.00 \times 10^{-6})(3.00 \times 10^{-6})}{(0.100)(0.300)} = \boxed{0.899\ \text{N}}, \tag{15.4.3}$$

directed away from the rod. It is tempting to treat the rod as a point charge at its center, $0.200\ \text{m}$ from q. That would give $0.674\ \text{N}$, which is 25% too small. The near half of the rod gains more force by being close than the far half loses by being far away, because of the steep $1/x^2$ dependence.

Example 15.8: A point charge on the axis of a charged ring

A thin ring of radius a carries a total charge Q spread uniformly around it. A point charge q sits on the ring's axis, a distance x from its center, as shown. (a) Find the electric force on q. (b) Evaluate it for $a = 0.100\ \text{m}$, $x = 0.100\ \text{m}$, $Q = +2.00\ \mu\text{C}$, and $q = +1.00\ \mu\text{C}$.



•Figure description

A ring of charge, seen in perspective as an ellipse, with its axis horizontal. A point charge q sits on the axis a distance x to the right of the ring's center. A charge element d q at the top of the ring, a radius a from the center, is joined to q by a line of length r, which makes angle theta with the axis. The force d F on q points along that line away from d q, and is split into a component along the axis and a component perpendicular to it, pointing down. The matching element at the bottom of the ring produces a force whose perpendicular component points up, so the perpendicular components cancel.

Solution

(a) Every element dq of the ring is the same distance from q:

$$r = \sqrt{x^2 + a^2}. \tag{15.4.4}$$

Each element exerts a force of magnitude $dF = k_e q\ dq/r^2$, directed along the line from the element to q. That line makes an angle θ with the axis, so $d\mathbf{F}$ has a component $dF\cos\theta$ along the axis and a component $dF\sin\theta$ perpendicular to it.

Now use symmetry. For every element dq there is an identical element on the opposite side of the ring, whose perpendicular component points the opposite way (its force is dashed in the sketch). When all the elements are added, the perpendicular components cancel in pairs, and only the axial components survive. With $\cos\theta = x/r$,

$$F = \int dF\cos\theta = \int k_e \frac{q\,dq}{r^2}\cdot\frac{x}{r} = \frac{k_e\,q\,x}{(x^2+a^2)^{3/2}}\int dq. \tag{15.4.5}$$

Because x and r are the same for every element, they come outside the integral, and $\int dq$ is simply the total charge Q on the ring. Therefore

$$F = \frac{k_e\,qQ\,x}{(x^2+a^2)^{3/2}}, \tag{15.8}$$

directed along the axis, away from the ring for like charges.

Checks. At the center of the ring, $x = 0$ and $F = 0$: the pushes from all sides cancel. Far away, $x \gg a$, the a^2 is negligible and $F \to k_eqQ/x^2$, the force of a point charge.

(b) With $x = a = 0.100\ \text{m}$, $x^2 + a^2 = 0.0200\ \text{m}^2$, and

$$F = \frac{(8.99\times10^9)(1.00\times10^{-6})(2.00\times10^{-6})(0.100)}{(0.0200)^{3/2}} = \boxed{0.636\,\text{N}}. \tag{15.4.6}$$

Equation (15.8) has an interesting form. The force is zero at the center of the ring and falls toward zero far away, so somewhere in

between it must reach a maximum. Setting $dF/dx = 0$ shows that the maximum is at $x = a/\sqrt{2}$. And because the force on a charge of *opposite* sign points back toward the center from either side, such a charge released near the center of the ring oscillates back and forth through it. Problem 15.25 shows that for small displacements this is simple harmonic motion (Chapter 28).

Note: Looking ahead: the charge q always factors out

Look back at equations (15.7) and (15.8). In both, the force on q is q multiplied by something that depends only on the source (its charge, shape, and size) and on *where* q is placed. The same is true of every force in this chapter, because Coulomb's law is proportional to q. That suggests dividing q out and describing what the source does to the space around it, before any test charge is placed there. That quantity, the force per unit charge, is the *electric field*, and it is the subject of Chapter 16. The rod and the ring will reappear there, along with a disk, an infinite plane, and a sphere.

Checkpoint 15.5

(a) In Example 15.8, the ring's charge is moved so that all of it sits on the upper half of the ring, still uniformly spread. Is the force on q still directed along the axis? Is its axial component larger, smaller, or the same as before? *(Hint: does moving an element around the ring change its distance from q?)* (b) Using equation (15.8), at what distance x, in terms of a, is the force on q greatest? Check the claim in the text that it is $a/\sqrt{2}$.

Answer

Answers to Checkpoints

Checkpoint 15.5. (a) No, the force is no longer purely along the axis. Each element still pushes q with a component along the axis, but the components perpendicular to the axis no longer cancel in pairs; the net force gains a component pointing away from the charged upper half, that is, downward. The axial component is *unchanged*: moving an element around the ring changes neither its distance r from q nor the angle its push makes with the axis, so each element contributes the same axial push as before. (b) Set $dF/dx = 0$. By the quotient rule the numerator of the derivative is $(x^2 + a^2)^{3/2} - x \cdot \frac{3}{2}(x^2 + a^2)^{1/2}(2x)$. Setting it to zero and dividing by $(x^2 + a^2)^{1/2}$ gives $x^2 + a^2 = 3x^2$, so $x^2 = a^2/2$ and $x = a/\sqrt{2}$, confirming the text.

This page titled 15.4: Electric Forces on Continuous Charge Distributions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.S: Electric Charges and Coulomb's Law (Summary)

Summary

- **Electric charge** comes in two kinds, positive and negative. Like charges repel and unlike charges attract. Protons carry charge $+e$, electrons $-e$, with $e = 1.602 \times 10^{-19}$ C, and a neutral object has equal amounts of each.
- Charge is **conserved**: the net charge of an isolated system never changes. Charging moves electrons from one object to another. Charge is also **quantized**: every free charge is a whole-number multiple of e, $q = ne$.
- In a **conductor**, free electrons move easily and charge spreads over the object. In an **insulator**, charge stays where it is placed. A conductor can be charged by **conduction** (contact, giving the same sign) or by **induction** (without contact, with a ground connection, giving the opposite sign). A nearby charge **polarizes** a neutral object, which is then attracted to it whatever the charge's sign.
- **Coulomb's law**: two point charges exert forces of magnitude $F_e = k_e|q_1||q_2|/r^2$ on each other, along the line joining them, with $k_e = 1/4\pi\epsilon_0 = 8.99 \times 10^9 \ \mathrm{N \cdot m^2/C^2}$. In vector form, $\mathbf{F}_{12} = k_e(q_1 q_2/r^2)\ \hat{r}_{21}$. Between fundamental particles, the electric force is 10^{36} to 10^{42} times stronger than gravity (about 10^{39} for an electron and a proton).
- By the **superposition principle**, the net force on a charge is the vector sum of the forces exerted by each of the other charges, each computed as though the others were absent. No arrangement of fixed charges can hold a charge in stable equilibrium (Earnshaw's theorem).
- A **continuous charge distribution** is described by a linear (λ), surface (σ), or volume (ρ) charge density. The force it exerts on a point charge is $\mathbf{F} = k_e q \int (dq/r^2)\ \hat{r}$, evaluated component by component, using symmetry to discard components that cancel.

Key Equations

Quantization of charge

$$q = ne, \qquad e = 1.602 \times 10^{-19}\,\mathrm{C} \tag{15.S.1}$$

Coulomb's law

$$F_e = k_e \frac{|q_1|\,|q_2|}{r^2}, \qquad k_e = \frac{1}{4\pi\epsilon_0} = 8.99 \times 10^9\,\mathrm{N \cdot m^2/C^2} \tag{15.S.2}$$

Coulomb's law, vector form

$$\mathbf{F}_{12} = k_e \frac{q_1 q_2}{r^2}\,\hat{r}_{21} \tag{15.S.3}$$

Superposition of electric forces

$$\mathbf{F}_1 = \sum_{j=2}^{N} \mathbf{F}_{1j} \tag{15.S.4}$$

Uniform charge densities

$$\lambda = \frac{Q}{L}, \qquad \sigma = \frac{Q}{A}, \qquad \rho = \frac{Q}{V} \tag{15.S.5}$$

Force from a continuous charge distribution

$$\mathbf{F} = k_e\, q \int \frac{dq}{r^2}\,\hat{r}, \qquad dq = \lambda\, d\ell, \quad \sigma\, dA, \text{ or } \rho\, dV \tag{15.S.6}$$

This page titled 15.S: Electric Charges and Coulomb's Law (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

15.E: Electric Charges and Coulomb's Law (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $e = 1.602 \times 10^{-19}$ C, $k_e = 8.99 \times 10^{9}$ N·m^2/C^2, $m_e = 9.11 \times 10^{-31}$ kg, and $m_p = 1.673 \times 10^{-27}$ kg.

Section 15.1 — Electric Charge

15.1 ★ A plastic comb run through dry hair acquires a charge of -8.0 nC.

a. Did the comb gain or lose electrons? How many?
b. What is the charge on the hair, and which law tells you?

15.2 ★ *[Conceptual.]* In repeating Millikan's experiment, a student reports oil-drop charges of 3.20×10^{-19} C, 8.01×10^{-19} C, and 5.60×10^{-19} C. Which of these values cannot be correct, and why?

15.3 ★★ *[Real-world: coins.]* A copper coin has a mass of 3.11 g. Copper has atomic number 29 (each neutral atom has 29 electrons) and a molar mass of 63.5 g/mol.

a. How many electrons does the coin contain?
b. What is their total charge?
c. If one electron in every billion were removed, what would the coin's net charge be?

15.4 ★ *[Conceptual.]* Explain why a metal rod held in your bare hand cannot be given a lasting charge by rubbing it with fur, but the same rod can be charged if it is held by a plastic handle.

15.5 ★★ *[Conceptual.]* Two identical neutral metal spheres, A and B, sit on insulating stands and touch each other. A negatively charged rod is brought near A without touching it. With the rod held in place, the spheres are pulled apart, and then the rod is removed.

a. What is the sign of the final charge on each sphere?
b. How do the magnitudes of the two charges compare? Which law guarantees this?
c. Would the result change if the rod were removed *before* the spheres were separated?

15.6 ★★ *[Conceptual.]* A balloon rubbed on your hair sticks to a neutral wall. (a) Explain why, with a sketch showing the charges in the wall. (b) Would a *positively* charged balloon also stick? (c) Why does the balloon eventually fall off, especially on a humid day?

Section 15.2 — Coulomb's Law

15.7 ★ A charge of $+3.0$ μC and a charge of -5.0 μC are 0.20 m apart. Find the magnitude of the force each exerts on the other, and state whether it is attractive or repulsive.

15.8 ★ *[Conceptual.]* Two point charges attract each other with a force F. By what factor does the force change if

a. the distance between them is halved?
b. one charge is tripled and the other is halved?
c. both charges are doubled and the distance is doubled?

15.9 ★★ Two identical small conducting spheres carry charges of $+6.0$ μC and -2.0 μC, with centers 0.30 m apart.

a. Find the force between them, and state whether it is attractive or repulsive.
b. The spheres are touched together and then returned to the same separation. What charge does each now carry, and what is the new force?

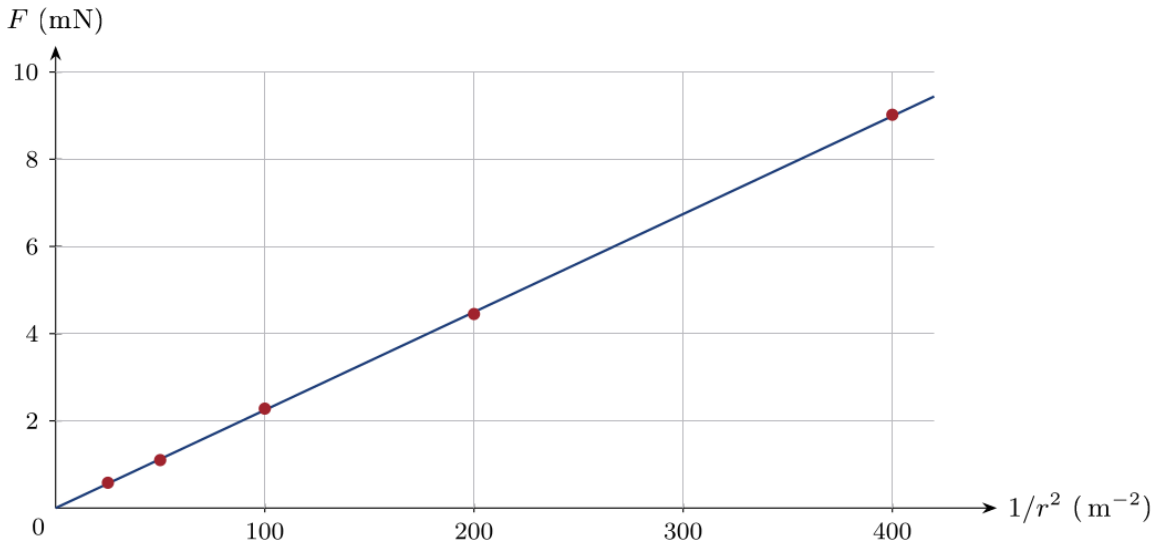
15.10 ★★ *[Real-world: atomic physics.]* In Rutherford's scattering experiment, an alpha particle (charge $+2e$, mass 6.64

$\times 10^{-27}$ kg) approaches a gold nucleus (charge $+79e$). At the point of closest approach the two are 3.0×10^{-14} m apart.

a. Find the electric force on the alpha particle there.
b. Find the alpha particle's acceleration, and compare it with g.

15.11 ★★★ Two tiny pith balls, each of mass 0.10 g, hang from threads 0.80 m long attached at the same point. After being given equal charges, they come to rest 5.0 cm apart. Find the charge on each ball.

15.12 ★★ *[Graph-reading.]* Two small spheres carrying equal charges q are placed at several separations r, and the force between them is measured. The graph plots the force F against $1/r^2$.



•Figure description
A graph of force in millinewtons on the vertical axis, from 0 to 10, against one over r squared in inverse square meters on the horizontal axis, from 0 to 400. Five measured points lie close to a straight line through the origin: about 0.58 millinewtons at 25, 1.10 at 50, 2.28 at 100, 4.45 at 200, and 9.02 at 400.

a. Explain why a straight line through the origin on this graph confirms the inverse-square law.
b. Find the slope of the line, in $\text{N} \cdot \text{m}^2$.
c. Use the slope to find q.

Section 15.3 — Discrete Charge Arrangements and the Superposition Principle

15.13 ★ Three point charges lie on the x-axis: $q_1 = +3.0$ μC at $x = 0$, $q_2 = -3.0$ μC at $x = 0.20$ m, and $q_3 = +4.0$ μC at $x = 0.50$ m. Find the magnitude and direction of the net force on q_2.

15.14 ★★ A charge of $+2.0$ μC is at the origin and a charge of -8.0 μC is at $x = 0.30$ m.

a. Explain why a third charge placed anywhere *between* them cannot feel zero net force.
b. Find the point on the x-axis where the net force on a third charge is zero.

15.15 ★★ Three charges sit at the corners of an equilateral triangle with sides of 0.15 m. Two of them are $+2.0$ μC and the third is -2.0 μC. Find the magnitude and direction of the net force on the negative charge.

15.16 ★★ Four equal charges $q = +1.0$ μC sit at the corners of a square with sides $a = 0.10$ m.

a. Show that the net force on any one of them points along the diagonal, away from the center of the square.
b. Show that its magnitude is $(k_e q^2/a^2)\left(\sqrt{2} + \frac{1}{2}\right)$, and evaluate it.

15.17 ★★★ Two equal charges $+Q$ are fixed at $(0,\ a)$ and $(0,\ -a)$. A charge $+q$ is placed on the x-axis at distance x from the origin.

a. Show that the net force on q is $F = 2k_eqQx/(x^2 + a^2)^{3/2}$, directed along the x-axis.
b. At what value of x is this force greatest?
c. Compare your result with equation (15.8), and explain why the two have the same form.

Section 15.4 — Electric Forces on Continuous Charge Distributions

15.18 ★ Find the charge density in each case.

a. A wire $1.5\ \text{m}$ long carries $6.0\ \text{nC}$ spread uniformly along its length.
b. A thin disk of radius $0.10\ \text{m}$ carries $4.0\ \text{nC}$ spread uniformly over one face.
c. A solid sphere of radius $5.0\ \text{cm}$ carries $2.0\ \text{nC}$ spread uniformly through its volume.

15.19 ★★ A thin rod $0.50\ \text{m}$ long carries a charge of $+4.0\ \mu\text{C}$ spread uniformly. A point charge of $-2.0\ \mu\text{C}$ sits on the rod's axis, $0.10\ \text{m}$ beyond one end.

a. Find the magnitude and direction of the force on the point charge.
b. By what percentage would you be wrong if you treated the rod as a point charge at its center?

15.20 ★★ A thin ring of radius $5.0\ \text{cm}$ carries $+3.0\ \mu\text{C}$ spread uniformly. Find the force on a $+0.50\ \mu\text{C}$ point charge on the ring's axis $12\ \text{cm}$ from its center.

15.21 ★★★ A thin rod bent into a semicircle of radius R carries a total charge Q spread uniformly. Show that the force on a point charge q at the center of the semicircle has magnitude $2k_eqQ/(\pi R^2)$, and find its direction. *(Hint: measure each element's position by its angle ϕ, so $d\ell = R\ d\phi$, and use symmetry to decide which component survives.)*

15.22 ★★★ A thin rod lies along the x-axis from $x = 0$ to $x = L$. Its charge is *not* uniform: the linear charge density is $\lambda = \lambda_0 x/L$, where λ_0 is a constant.

a. Find the total charge on the rod.
b. A point charge q is placed on the axis at $x = -d$. Show that the force on it has magnitude

$$F = \frac{k_eq\lambda_0}{L}\left[\ln\left(\frac{d+L}{d}\right) - \frac{L}{d+L}\right]. \tag{15.E.1}$$

15.23 ★★★ Two identical small conducting spheres, with centers $0.500\ \text{m}$ apart, attract each other with a force of $0.108\ \text{N}$. They are connected briefly by a thin conducting wire, which is then removed, and afterward they repel each other with a force of $0.0360\ \text{N}$. What were the initial charges on the spheres? *(There are two sets of answers, differing only in sign.)*

15.24 ★★★ *[Real-world: atomic physics.]* In the Bohr model of hydrogen (Chapter 41), the electron moves in a circular orbit of radius $5.29 \times 10^{-11}\ \text{m}$ around the proton, held by the Coulomb attraction.

a. Find the electron's orbital speed, and express it as a fraction of the speed of light.
b. Find the time for one orbit.
c. Explain why the gravitational attraction between the proton and electron can be left out of this calculation entirely.

15.25 ★★★ A thin ring of radius a carries a positive charge Q spread uniformly. A particle of mass m and negative charge $-q$ is placed on the ring's axis a small distance $x \ll a$ from the center and released.

a. Use equation (15.8) to show that, for $x \ll a$, the force on the particle is approximately $F_x = -(k_eqQ/a^3)\ x$.
b. Explain why the particle undergoes simple harmonic motion, and show that its angular frequency is $\omega = \sqrt{k_eqQ/(ma^3)}$.
c. Evaluate the period for an electron and a ring with $a = 5.0\ \text{cm}$ and $Q = +10\ \text{nC}$.
d. Why is this not a counterexample to Earnshaw's theorem? *(Hint: what happens if the particle is displaced slightly off the axis?)*

This page titled 15.E: Electric Charges and Coulomb's Law (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16: ELECTRIC FIELDS

Coulomb's law tells us the force between two charges, but it is silent on a question that troubled Newton about gravity and troubled Faraday about electricity: *how* does one charge know the other is there? Two charges a meter apart push on each other across empty space, with nothing visible passing between them. Newton called action at a distance "so great an absurdity" that no thoughtful person could accept it, yet he had no alternative to offer.

Michael Faraday did. He pictured every charge as filling the space around it with an invisible influence, a *field*, and each other charge as responding to the field at its own location rather than to the distant charge directly. For static charges this is only a change of viewpoint, and the forces come out the same either way. But it turns out to be far more than bookkeeping. When a charge moves, the change in its field spreads outward at a finite speed: the speed of light. Further, a field disturbance racing through empty space, leaving nothing behind, is exactly what a light wave is (Chapter 27). Fields carry energy and momentum of their own. They are as real as the charges that make them.

This chapter develops the electric field. We define it, find it for point charges and for the electric dipole, and then calculate it for the continuous distributions that appear throughout the rest of electricity and magnetism: lines, disks, rings, planes, and spheres. Finally we turn the question around and ask what a known field does to a charge, or to a dipole, placed in it.

Learning Objectives

- Define the electric field as force per unit charge, and draw and interpret electric field lines.
- Calculate the electric field of point charges by superposition, and the field of an electric dipole.
- Integrate over a continuous charge distribution to find the field of a line of charge, a disk, a ring, an infinite plane, and a sphere, and check each result in its limits.
- Analyze the motion of a charged particle in a uniform electric field.
- Find the torque on, and the potential energy of, an electric dipole in an electric field.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

16.1: The Electric Field
16.2: The Electric Field of Point Charges
16.3: The Electric Field of Continuous Charge Distributions
16.4: A Point Charge in an Electric Field
16.5: A Dipole in an Electric Field
16.S: Electric Fields (Summary)
16.E: Electric Fields (Problems)

This page titled 16: Electric Fields is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.1: The Electric Field

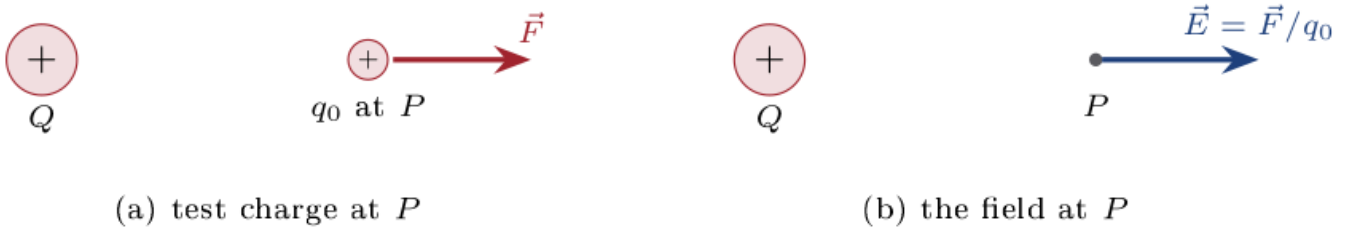


Figure 16.1. (a) A test charge q_0 at point P feels a force $\mathbf{F}$ from the source charge Q. (b) The electric field at P is $\mathbf{E} = \mathbf{F}/q_0$. It depends only on the source and the location, and it is there whether or not a test charge is present to measure it.

•Figure description

Two panels. In panel a, a positive source charge Q sits at the left and a small positive test charge q sub zero sits at point P to its right; an arrow F on the test charge points directly away from Q. In panel b, the test charge has been removed; at the same point P an arrow labeled E points in the same direction, representing the electric field that exists at P whether or not a test charge is there.

Suppose a charge Q, or any arrangement of charges, sits somewhere in space. We call it the **source** of the field. To explore the space around it, we carry a small positive **test charge** q_0 from point to point and measure the electric force $\mathbf{F}$ on it at each point P (Figure 16.1). By Coulomb's law, that force is proportional to q_0: double the test charge and the force doubles. The ratio $\mathbf{F}/q_0$ is therefore independent of the test charge. It depends only on the source and on the location of P, and it is what we call the **electric field** at P:

Key Equation: Electric field

$$\mathbf{E} = \frac{\mathbf{F}}{q_0}. \tag{16.1}$$

The electric field is a vector. Its direction is the direction of the force on a *positive* test charge, and its SI unit is the newton per coulomb ($\mathrm{N/C}$). In Chapter 18 you will meet the volt, and see that $\mathrm{N/C}$ is the same as the volt per meter ($\mathrm{V/m}$), the unit most often used in practice.

Two cautions about the test charge. First, it must be small. A large test charge would push the source charges around (on a conductor, say) and so change the very field it was meant to measure. Strictly, $\mathbf{E}$ is the limit of $\mathbf{F}/q_0$ as $q_0 \to 0$. Second, the field at P exists whether or not any test charge is there. The test charge is only a way of detecting it, much as a thermometer detects a temperature that was already there.

Once the field is known, the force on *any* charge q placed at P follows immediately from equation (16.1):

Key Equation: Force on a charge in an electric field

$$\mathbf{F} = q\mathbf{E}. \tag{16.2}$$

If q is positive, the force points along $\mathbf{E}$. If q is negative, the force points opposite to $\mathbf{E}$. The field has split the electric interaction into two steps: *charges create fields*, and *fields exert forces on charges*. The first step is the subject of Sections 16.2–16.3, the second of Sections 16.4 and 16.5.

The gravitational field works the same way. In Chapter 13 the free-fall acceleration g was the gravitational force per unit mass, $\mathrm{g} = \mathbf{F}_g/m$, and the weight of any object was $m\mathrm{g}$. The electric field $\mathbf{E}$ plays the role of g, with charge in place of mass.

Table 16.1 gives a sense of scale. The fields in it range over more than twenty orders of magnitude.

Table 16.1. Approximate magnitudes of some electric fields.

Location	Field magnitude ($\mathrm{N/C}$)
Inside a copper wire carrying household current	10^{-2}
Near the ground in fair weather	10^{2}
Sunlight at the top of the atmosphere (rms)	7×10^{2}
Near the ground beneath a thundercloud	10^{4}
Photocopier drum	10^{5}
At which dry air breaks down and sparks	3×10^{6}
At the electron's orbit in a hydrogen atom	5×10^{11}
At the surface of a uranium nucleus	2×10^{21}

Example 16.1: Force on an electron from a measured field

A test charge $q_0 = +3.0\ \mathrm{nC}$ placed at a point P feels a force of $6.0 \times 10^{-4}\ \mathrm{N}$ directed due east. (a) What is the electric field at P? (b) The test charge is removed and an electron is placed at P. What force does it feel, and what is its acceleration? (Take $m_e = 9.11 \times 10^{-31}\ \mathrm{kg}$.)

Solution

(a) From equation (16.1),

$$E = \frac{F}{q_0} = \frac{6.0 \times 10^{-4}\,\mathrm{N}}{3.0 \times 10^{-9}\,\mathrm{C}} = \boxed{2.0 \times 10^{5}\,\mathrm{N/C}}, \tag{16.1.1}$$

directed due east, the direction of the force on the positive test charge.

(b) The field is unchanged by removing the test charge. By equation (16.2), the force on the electron is

$$F = |q|E = (1.602 \times 10^{-19}\,\mathrm{C})(2.0 \times 10^{5}\,\mathrm{N/C}) = \boxed{3.2 \times 10^{-14}\,\mathrm{N}}, \tag{16.1.2}$$

directed due *west*, because the electron's charge is negative. Its acceleration is

$$a = \frac{F}{m_e} = \frac{3.2 \times 10^{-14}\,\mathrm{N}}{9.11 \times 10^{-31}\,\mathrm{kg}} \approx \boxed{3.5 \times 10^{16}\,\mathrm{m/s^2}}, \tag{16.1.3}$$

also westward. That is some 10^{15} times g. A tiny force acting on a tiny mass gives an enormous acceleration, which is why electrons in fields reach high speeds in very short distances (Section 16.4).

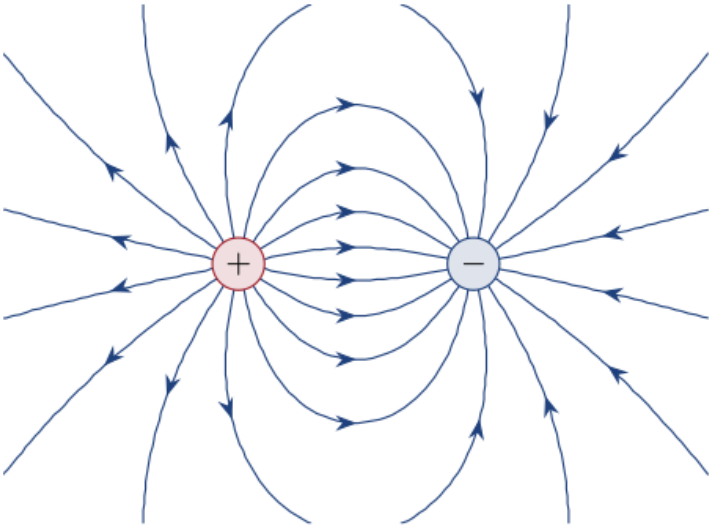


Figure 16.2. Field lines of an electric dipole, a positive and a negative charge of equal size. Every line that leaves the positive charge ends on the negative charge (some after a wide loop that runs off this drawing).

•Figure description

Computed field lines of an electric dipole: a positive charge on the left and an equal negative charge on the right. Lines leave the positive charge in all directions and curve around to end on the negative charge. The lines are densest in the region between the two charges, where the field points from the positive toward the negative charge, and they bulge outward above and below in wide arcs.

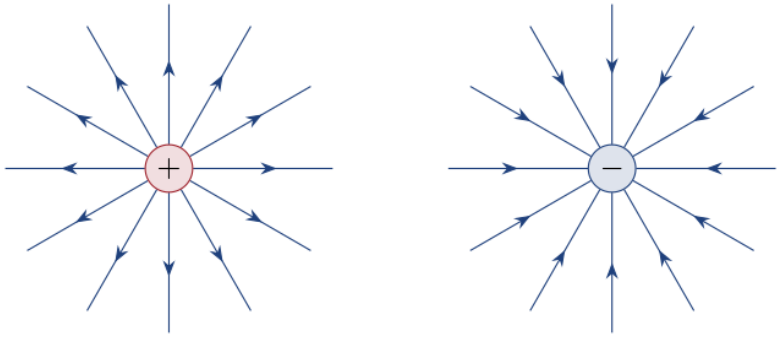


Figure 16.3. Field lines of a positive point charge (left) point radially outward. Those of a negative point charge (right) point radially inward. The lines crowd together near the charge, where the field is strong, and spread out farther away.

•Figure description

Two point charges, each surrounded by twelve straight, evenly spaced field lines. The lines around the positive charge on the left carry arrows pointing outward, away from the charge. The lines around the negative charge on the right carry arrows pointing inward, toward the charge. In both drawings the lines are crowded near the charge and spread apart farther away.

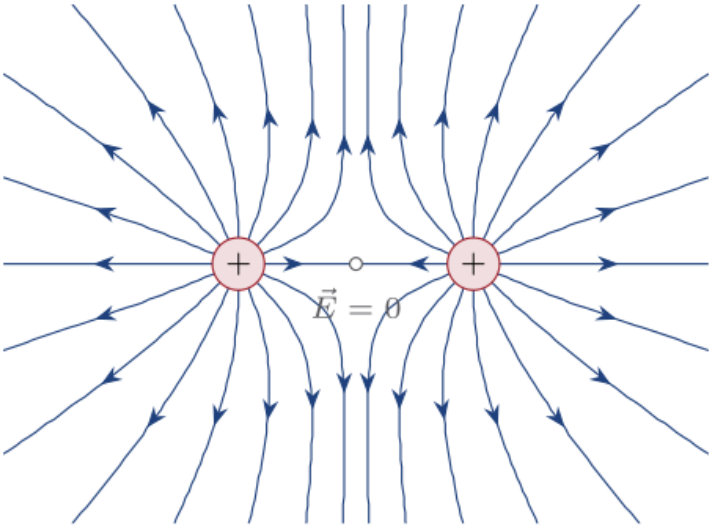


Figure 16.4. Field lines of two equal positive charges. The lines from each charge bend away from the other, and the field is zero at the point midway between them. Far away, the pattern approaches that of a single charge $2q$.

•Figure description

Computed field lines of two equal positive charges side by side. Lines leave each charge and bend away from the other charge. Midway between the two charges there is a point where no line passes, because the two fields cancel there. Farther out, the lines stream outward in all directions.

Electric field lines. Faraday visualized the field with **electric field lines**, also called lines of force. They are drawn according to three rules:

1. At every point, the field line through that point is tangent to $\mathbf{E}$ there, and an arrow on the line shows the direction of $\mathbf{E}$.
2. Field lines begin on positive charges and end on negative charges (or run off to infinity). The number of lines leaving or entering a charge is proportional to the magnitude of the charge.
3. The lines are closest together where the field is strongest. More precisely, the number of lines crossing a small area perpendicular to them is proportional to the magnitude of $\mathbf{E}$ there. In other words, the density of lines is proportional to the magnitude of $\mathbf{E}$.

It follows that two field lines can never cross. If they did, the field at the crossing point would have two directions at once.

For a single point charge the lines are straight and radial: outward from a positive charge and inward toward a negative one (Figure 16.3). The lines spread apart with distance. The same number of lines crosses every sphere centered on the charge, and the area of the sphere grows as r^2, so the density of lines falls as $1/r^2$, exactly as the field does. That the field-line picture works at all is a consequence of the inverse-square law, and in Chapter 17 this idea of counting lines through a surface becomes Gauss's law.

Figures 16.2 and 16.4 show the field lines for two pairs of equal charges. Between a positive and a negative charge, the lines leave one charge and curve around to end on the other, and they are densest in the region between the two. Between two positive charges, the lines from each charge bend away from the other. Midway between the charges the two fields cancel, and there no field line passes at all. Far from either pair, the field looks quite different in the two cases. The opposite charges' lines mostly close on each other, so their field fades quickly with distance. The like charges' lines stream outward in every direction, as though from a single charge $2q$ at the center.

Note: Field lines are not trajectories

It is tempting to think that a charge released at rest will travel along a field line. In general it will not. The field line gives the direction of the *force*, and so of the *acceleration*, but a moving particle's velocity need not point along its acceleration. A projectile's path is not a line of the gravitational field either. A charge follows a field line only when the lines are straight, as near a point charge or in a uniform field, and it starts from rest.

Checkpoint 16.1

(a) Explain, using the rules for field lines, why two field lines can never cross. (b) In Figure 16.4, where is the field weakest, and how can you tell from the drawing alone? (c) A student draws 12 lines leaving a charge q_1 and 36 lines entering a charge q_2. What can you conclude about the two charges?

Answer

Answers to Checkpoints

Checkpoint 16.1. (a) A field line is drawn tangent to the field, and at each point the field has exactly one direction. If two lines crossed, the field at the crossing would have to point along both at once, which is impossible. (b) Where the lines are sparsest. For two like charges that is the region midway between them, where the lines from the two charges bend away from each other and leave a gap; at the exact midpoint the field is zero. Line density is proportional to field strength, so the drawing alone shows it. (c) Lines leave positive charges and end on negative ones, and the number of lines is proportional to the size of the charge. So q_1 is positive, q_2 is negative, and $|q_2| = 3|q_1|$.

This page titled 16.1: The Electric Field is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.2: The Electric Field of Point Charges

The field of a single point charge q follows directly from Coulomb's law. A test charge q_0 at distance r feels a force $\mathbf{F} = k_e(qq_0/r^2)\,\hat{r}$, where $\hat{r}$ is the unit vector pointing from q toward the test charge (equation (15.3)). Dividing by q_0,

Key Equation: Electric field of a point charge

$$\mathbf{E} = k_e \frac{q}{r^2}\,\hat{r}. \tag{16.3}$$

For positive q the field points away from the charge, and for negative q it points toward the charge. Its magnitude, $E = k_e|q|/r^2$, falls off as the inverse square of the distance. Notice that the test charge has disappeared from the result, as it must.

The electric force obeys the superposition principle (Section 15.3), and so does the field. The force on a test charge is the vector sum of the forces from all the source charges. Dividing every term by q_0, the field is the vector sum of the fields of the individual charges:

Key Equation: Superposition of electric fields

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \cdots = \sum_i k_e \frac{q_i}{r_i^2}\,\hat{r}_i. \tag{16.4}$$

Here r_i is the distance from charge q_i to the point where the field is wanted, and $\hat{r}_i$ points from q_i toward that point. The strategy of Section 15.3 applies without change: find each field's magnitude, decide its direction (away from positive charges, toward negative ones), resolve into components, and add.

Example 16.2: The field of a point charge

What is the electric field $0.30\ \text{m}$ from a point charge of $+5.0\ \mu\text{C}$? What force would a charge of $-2.0\ \text{nC}$ feel if placed there?

Solution

From equation (16.3),

$$E = k_e\frac{|q|}{r^2} = \frac{(8.99\times10^9)(5.0\times10^{-6})}{(0.30)^2} = \boxed{5.0\times10^5\ \text{N/C}}, \tag{16.2.1}$$

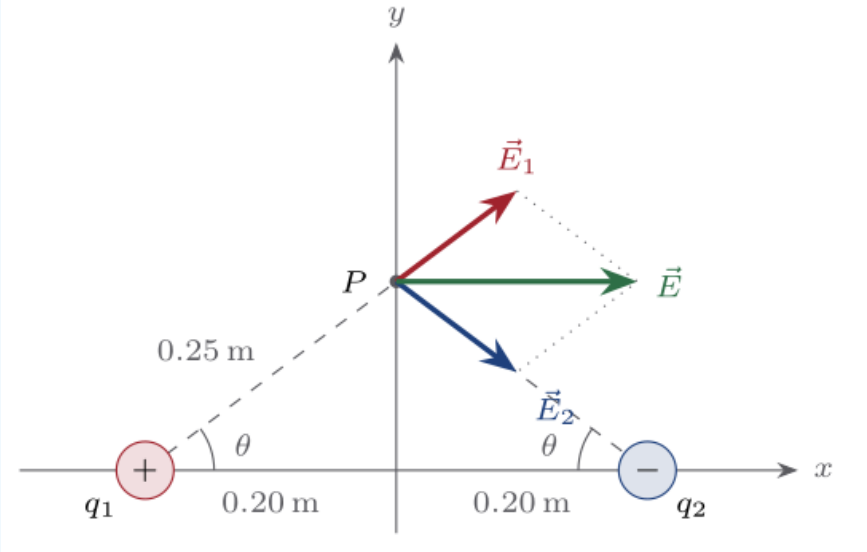
directed radially away from the charge. The force on the $-2.0\ \text{nC}$ charge is

$$F = |q|E = (2.0\times10^{-9})(5.0\times10^5) = \boxed{1.0\times10^{-3}\ \text{N}}, \tag{16.2.2}$$

directed toward the $5.0\ \mu\text{C}$ charge, opposite to $\mathbf{E}$. Coulomb's law gives the same answer directly, but the field has done something useful: once E is known at a point, the force on any charge placed there takes one multiplication.

Example 16.3: The field of two opposite charges

A charge $q_1 = +8.0\ \text{nC}$ sits at $(-0.20\ \text{m},\ 0)$ and a charge $q_2 = -8.0\ \text{nC}$ sits at $(+0.20\ \text{m},\ 0)$. Find the electric field at the point P at $(0,\ 0.15\ \text{m})$.



•Figure description

A positive charge q sub 1 on the left and a negative charge q sub 2 on the right lie on the x-axis, 0.40 meters apart. Point P sits on the y-axis 0.15 meters above the midpoint. The dashed lines from each charge to P are 0.25 meters long and make angle theta with the x-axis. At P, the field E sub 1 points up and to the right, away from q sub 1, and the field E sub 2 points down and to the right, toward q sub 2. Their sum E points horizontally to the right, parallel to the line from the positive to the negative charge.

Solution

Each charge is $\sqrt{0.20^2 + 0.15^2} = 0.25\ \text{m}$ from P, and the two charges have equal magnitudes, so the two fields have equal magnitudes:

$$E_1 = E_2 = k_e \frac{|q|}{r^2} = \frac{(8.99 \times 10^9)(8.0 \times 10^{-9})}{(0.25)^2} = 1.15 \times 10^3\ \text{N/C}. \tag{16.2.3}$$

$\mathbf{E}_1$ points away from the positive charge q_1, along the dashed line from q_1 through P, and $\mathbf{E}_2$ points toward the negative charge q_2, along the dashed line from P to q_2. Each line makes the same angle θ with the x-axis. From the right triangle formed by the charge, the origin, and P,

$$\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{0.20\,\text{m}}{0.25\,\text{m}} = 0.80, \qquad \sin\theta = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{0.15\,\text{m}}{0.25\,\text{m}} = 0.60. \tag{16.2.4}$$

(The angle itself is $\theta = 36.9°$, but we need only its cosine and sine.) $\mathbf{E}_1$ points up and to the right and $\mathbf{E}_2$ points down and to the right, so their components are

$$\begin{aligned} E_{1x} &= E_1 \cos\theta, \qquad & E_{1y} &= +E_1 \sin\theta, \\ E_{2x} &= E_2 \cos\theta, \qquad & E_{2y} &= -E_2 \sin\theta. \end{aligned} \tag{16.2.5}$$

Because $E_1 = E_2$, the y-components cancel, and the x-components are equal and add:

$$E_y = E_1 \sin\theta - E_2 \sin\theta = 0, \qquad E_x = E_1 \cos\theta + E_2 \cos\theta = 2E_1 \cos\theta. \tag{16.2.6}$$

Therefore

$$\mathbf{E} = 2(1.15 \times 10^3\ \text{N/C})(0.80)\,\hat{\imath} = \boxed{(1.84 \times 10^3\ \text{N/C})\,\hat{\imath}}. \tag{16.2.7}$$

The field at P points in the $+x$ direction, parallel to the line from the positive charge to the negative one. This arrangement of two equal and opposite charges is important enough to have a name, and it is the subject of the next subsection.

Checkpoint 16.2

A charge of $+9.0\ \mathrm{nC}$ sits at $x = 0$ and a charge of $+4.0\ \mathrm{nC}$ sits at $x = 0.50\ \mathrm{m}$. (a) Explain why the field can be zero only at some point between them. (b) Find that point. (c) If the second charge were $-4.0\ \mathrm{nC}$ instead, in which region would the zero-field point lie?

Answer

The Electric Field of an Electric Dipole

A pair of equal and opposite charges, $+q$ and $-q$, separated by a distance d, is called an **electric dipole**. Dipoles are everywhere. Many molecules, water among them, have their electron clouds shifted toward one end, so that one end is slightly negative and the other slightly positive. And in Chapter 15 we saw that a nearby charge polarizes any neutral atom or molecule, turning it into a small dipole. Radio and TV antennas are dipoles whose charges oscillate back and forth.

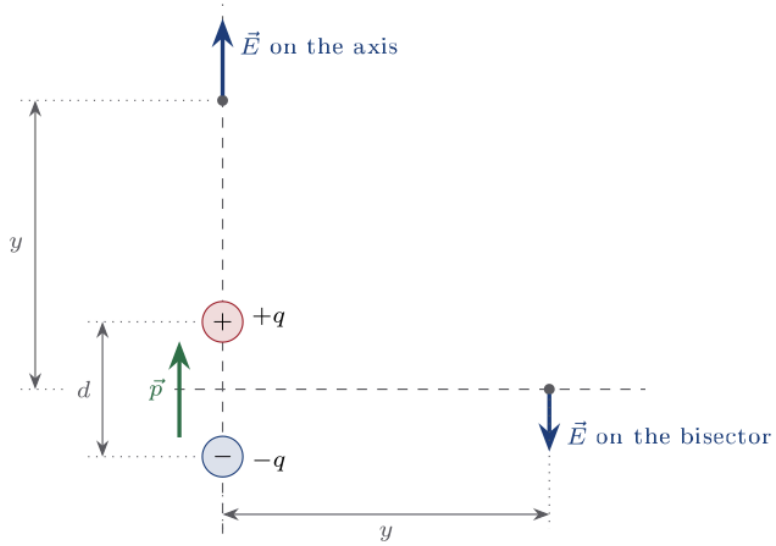


Figure 16.5. An electric dipole. The dipole moment $\mathbf{p}$ points from $-q$ to $+q$. On the axis, the field points along $\mathbf{p}$. On the perpendicular bisector, it points opposite to $\mathbf{p}$ and is half as strong at the same large distance.

•Figure description

An electric dipole drawn vertically, with the positive charge on top and the negative charge below, separated by a distance d. The dipole moment vector p points upward from the negative to the positive charge. A point on the axis, a distance y above the center, has a field E pointing upward, parallel to p. A point on the perpendicular bisector, a distance y to the right of the center, has a field E pointing downward, opposite to p.

A dipole is described by its **electric dipole moment**, a vector $\mathbf{p}$ that points from the negative charge to the positive charge (Figure 16.5) and has magnitude

Key Equation: Electric dipole moment

$$p = qd. \tag{16.5}$$

Its SI unit is the coulomb-meter ($\mathrm{C \cdot m}$). As we will see, far from a dipole its field depends on q and d only through the product p . A large charge with a small separation and a small charge with a large separation produce the same distant field, provided qd is the same.

On the perpendicular bisector. Example 16.3 already found the field at a point on the perpendicular bisector of a dipole. Repeating that calculation with symbols, for a point a distance y from the dipole's center, each charge is a distance $\sqrt{y^2 + (d/2)^2}$ away. The components along the bisector cancel, and the components parallel to the dipole axis add, pointing from $+q$ toward $-q$, opposite to $\mathbf{p}$. Each contributes $k_e q/(y^2 + d^2/4)$ times the fraction $(d/2)/\sqrt{y^2 + d^2/4}$, so

$$E_\perp = \frac{k_e q d}{(y^2 + d^2/4)^{3/2}} = \frac{k_e p}{(y^2 + d^2/4)^{3/2}}. \tag{16.6}$$

Check it against Example 16.3: with $p = (8.0 \times 10^{-9})(0.40) = 3.2 \times 10^{-9}\ \text{C}\cdot\text{m}$ and $y = 0.15\ \text{m}$, equation (16.6) gives $1.84 \times 10^3\ \text{N/C}$, as before.

On the axis. At a point on the axis a distance y from the center, on the side of the positive charge, the positive charge is a distance $y - d/2$ away and the negative charge is $y + d/2$ away. Both fields lie along the axis, the positive charge's pointing away from the dipole and the negative charge's pointing back toward it. The net field is

$$E_\parallel = k_e q \left[\frac{1}{(y - d/2)^2} - \frac{1}{(y + d/2)^2}\right] = \frac{2k_e q\, d\, y}{(y^2 - d^2/4)^2}, \tag{16.2.8}$$

directed along $\mathbf{p}$. (Combine the fractions over a common denominator to check the second step.)

Far from the dipole. Usually we want the field at distances much greater than the size of the dipole, $y \gg d$. Then the $d^2/4$ terms in the denominators are negligible, and the two results become

Key Equation: Field of a dipole at large distance

$$E_\parallel \approx \frac{2k_e p}{y^3} \quad \text{(on the axis)}, \qquad E_\perp \approx \frac{k_e p}{y^3} \quad \text{(on the bisector)}. \tag{16.7}$$

Two features stand out. The field depends only on p, as promised. And it falls off as $1/y^3$, one power faster than the field of a point charge. From far away the two charges nearly cancel, and what survives is the small effect of their separation. This is the quantitative version of what Figure 16.2 shows: a dipole's field lines close on themselves, and little of the field reaches far from the dipole.

Example 16.4: The field of a water molecule

A water molecule has an electric dipole moment of $p = 6.2 \times 10^{-30}\ \text{C}\cdot\text{m}$. Find the electric field it produces at a point on its axis $1.0\ \text{nm}$ away, and compare it with the field of a single elementary charge at the same distance.

Solution

The size of the molecule is about $0.1\ \text{nm}$, so $1.0\ \text{nm}$ is far enough for equation (16.7):

$$E_\parallel = \frac{2k_e p}{y^3} = \frac{2(8.99 \times 10^9)(6.2 \times 10^{-30})}{(1.0 \times 10^{-9})^3} \approx \boxed{1.1 \times 10^8\,\text{N/C}}. \tag{16.2.9}$$

A single charge e at the same distance would produce

$$E = \frac{k_e e}{y^2} = \frac{(8.99 \times 10^9)(1.602 \times 10^{-19})}{(1.0 \times 10^{-9})^2} = 1.4 \times 10^9\,\text{N/C}, \tag{16.2.10}$$

about 13 times more. The gap widens with distance: at $10\ \text{nm}$ the ion's field is 100 times smaller but the dipole's is 1000 times smaller. This is why the forces between neutral molecules are so much shorter-ranged than the forces between ions.

Checkpoint 16.3

At a distance r along its axis, a certain dipole produces a field of $800\ \text{N/C}$. (a) What field does it produce at a distance $2r$ along its axis? (b) At a distance $2r$ along its perpendicular bisector? (c) A point charge produces $800\ \text{N/C}$ at distance r. What does it produce at $2r$?

Answer

Answers to Checkpoints

Checkpoint 16.2. (a) Outside the pair, both fields point the same way (away from both positive charges), so they add and cannot cancel. Only between the charges do the two fields point in opposite directions. (b) Set the magnitudes equal at a distance x from the $9.0\ \mathrm{nC}$ charge: $k_e(9.0)/x^2 = k_e(4.0)/(0.50-x)^2$. Taking square roots, $3.0/x = 2.0/(0.50-x)$, so $1.5 - 3.0x = 2.0x$ and $x = 0.30\ \mathrm{m}$. The zero point is $0.30\ \mathrm{m}$ from the larger charge: closer to the *smaller* one, as it must be. (c) With opposite signs, the fields cancel only *outside* the pair, on the side of the smaller-magnitude charge: beyond $x = 0.50\ \mathrm{m}$. (Solving $3.0(x-0.50) = 2.0x$ puts it at $x = 1.5\ \mathrm{m}$.)

Checkpoint 16.3. Far from the dipole, its field falls off as $1/r^3$. (a) At $2r$ on the axis, $800/2^3 = 100\ \mathrm{N/C}$. (b) On the perpendicular bisector the field at a given distance is half the on-axis value, so at $2r$ it is $50\ \mathrm{N/C}$. (c) A point charge's field falls as $1/r^2$, so at $2r$ it is $800/2^2 = 200\ \mathrm{N/C}$. The dipole's field fades faster because, from far away, its two opposite charges increasingly cancel each other.

This page titled 16.2: The Electric Field of Point Charges is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.3: The Electric Field of Continuous Charge Distributions

In Section 15.4 we found the force that a continuous charge distribution exerts on a point charge q, and noticed that q always factors out. Dividing it out gives the field. Split the distribution into elements dq, each acting as a point charge with field $d\mathbf{E} = k_e\,(dq/r^2)\,\hat{r}$, and add them all up:

Key Equation: Electric field of a continuous charge distribution

$$\mathbf{E} = \int d\mathbf{E} = k_e \int \frac{dq}{r^2}\,\hat{r}. \tag{16.8}$$

Here r is the distance from the element dq to the point P where the field is wanted, and $\hat{r}$ points from dq toward P. The charge element is written with a charge density: $dq = \lambda\ d\ell$ along a line, $\sigma\ dA$ over a surface, or $\rho\ dV$ through a volume. The five-step procedure of Section 15.4 carries over unchanged: pick a typical element, express dq, r, and the direction of $d\mathbf{E}$ in terms of one coordinate, let symmetry cancel what it can, integrate the rest, and check the limits.

The next five subsections apply equation (16.8) to a line, a ring, a disk, a plane, and a sphere. Each shape is built on the ones before it, and every result must reduce to the field of a point charge when we are far away, a check we will make each time.

The Electric Field of a Line Charge

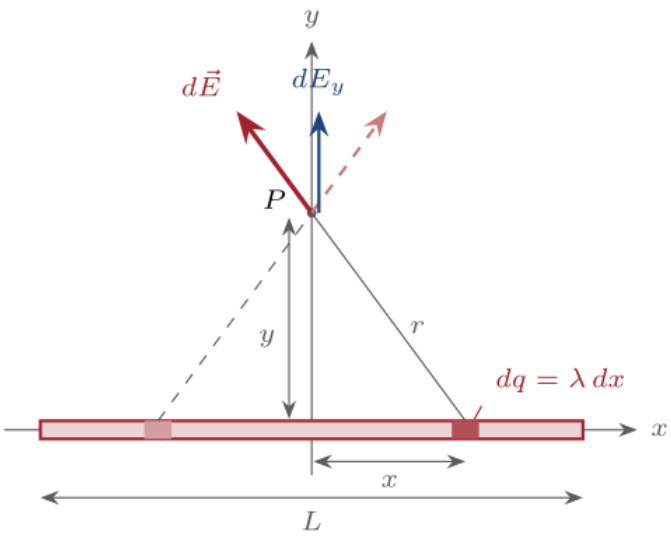


Figure 16.6. The field of a uniformly charged rod at a point P on its perpendicular bisector. The element dq at x produces $d\mathbf{E}$ (solid), and the mirror-image element at $-x$ produces the dashed field. Their components along the rod cancel, and their components dE_y perpendicular to the rod add.

•Figure description

A uniformly charged rod of length L lies along the x-axis, centered on the origin. Point P is on the y-axis a distance y above the rod's center. A small element d q equals lambda d x at position x to the right of center sends a field d E to P, pointing up and to the left, away from the element; it is split into a vertical component and a horizontal component. A matching element at minus x sends a field whose horizontal component points the other way, so the horizontal components cancel and only the vertical components add.

A thin rod of length L carries a charge Q spread uniformly, so its linear charge density is $\lambda = Q/L$. We want the field at a point P a distance y from the rod, on its perpendicular bisector (Figure 16.6).

Put the rod on the x-axis from $x = -L/2$ to $x = +L/2$, with P on the y-axis. An element of length dx at position x carries charge $dq = \lambda\ dx$ and is a distance $r = \sqrt{x^2 + y^2}$ from P. Its field $d\mathbf{E}$ points along the line from the element to P, and has components

$$dE_x = -dE\,\frac{x}{r}, \qquad dE_y = dE\,\frac{y}{r}, \qquad \text{where } dE = k_e\frac{\lambda\,dx}{r^2}. \tag{16.3.1}$$

Now symmetry. The element at $-x$ is the mirror image of the one at $+x$. Its y-component is the same, but its x-component is opposite. Adding the whole rod, the x-components cancel in pairs and $E_x = 0$. Only the y-components survive:

$$E_y = \int dE_y = k_e\lambda y\int_{-L/2}^{L/2}\frac{dx}{(x^2+y^2)^{3/2}}. \tag{16.3.2}$$

The integral is a standard one (differentiate the result to check it):

$$\int\frac{dx}{(x^2+y^2)^{3/2}} = \frac{x}{y^2\sqrt{x^2+y^2}}. \tag{16.3.3}$$

Evaluating between the limits and multiplying by $k_e\lambda y$,

$$E = \frac{k_e\lambda L}{y\sqrt{y^2+L^2/4}} = \frac{k_eQ}{y\sqrt{y^2+L^2/4}}, \tag{16.9}$$

directed perpendicular to the rod, away from it for positive Q.

Checks. Far from the rod, $y \gg L$, the square root becomes y and $E \to k_eQ/y^2$: from a distance, the rod looks like a point charge. The opposite limit is more interesting. Close to a long rod, $y \ll L$, the square root becomes $L/2$ and

$$E \approx \frac{k_e\lambda L}{y\,(L/2)} = \frac{2k_e\lambda}{y}. \tag{16.3.4}$$

This is the field of an **infinite line of charge**, and it is worth writing in terms of ϵ_0:

Key Equation: Field of an infinite line of charge

$$E = \frac{2k_e\lambda}{r} = \frac{\lambda}{2\pi\epsilon_0 r}. \tag{16.10}$$

(We have written r for the perpendicular distance from the line.) The field of a line falls off as $1/r$, more slowly than a point charge's $1/r^2$. By symmetry it points radially away from the line, so its field lines spread out in two dimensions rather than three: through a cylinder of radius r and area $2\pi rL$ around the line, not a sphere. Chapter 17 will derive equation (16.10) in two lines from exactly that observation.

Example 16.5: How long is "infinite"?

A rod $0.200\ \text{m}$ long carries $+5.00\ \text{nC}$ spread uniformly. Find the electric field at a point $5.00\ \text{cm}$ from the rod's center, on its perpendicular bisector, and compare it with the infinite-line and point-charge approximations.

Solution

From equation (16.9), with $y = 0.0500\ \text{m}$ and $L/2 = 0.100\ \text{m}$,

$$E = \frac{(8.99\times10^9)(5.00\times10^{-9})}{(0.0500)\sqrt{(0.0500)^2+(0.100)^2}} = \frac{44.95}{(0.0500)(0.1118)} = \boxed{8.04\times10^3\ \text{N/C}}, \tag{16.3.5}$$

directed away from the rod. The linear charge density is $\lambda = 2.50\times10^{-8}\ \text{C/m}$, so the infinite-line formula gives $2k_e\lambda/y = 8.99\times10^3\ \text{N/C}$, about 12% too high. Treating the rod as a point charge gives $k_eQ/y^2 = 1.80\times10^4\ \text{N/C}$, more than twice too high. At a distance of half the rod's half-length, the rod is already much better described as infinite than as a point. As a rule of thumb, the infinite-line formula is good whenever you are much closer to the line than to either of its ends.

The Electric Field of a Charged Ring

A thin ring of radius a carries a charge Q spread uniformly around it. We want the field at a point on its axis, a distance x from its center. Chapter 15 has already done the work. Example 15.8 found the force on a point charge q at exactly this point, using the symmetry of the ring to cancel every component perpendicular to the axis. Dividing equation (15.8) by q gives the field:

Key Equation: Field on the axis of a charged ring

$$E = \frac{k_e Q\, x}{(x^2 + a^2)^{3/2}}, \tag{16.11}$$

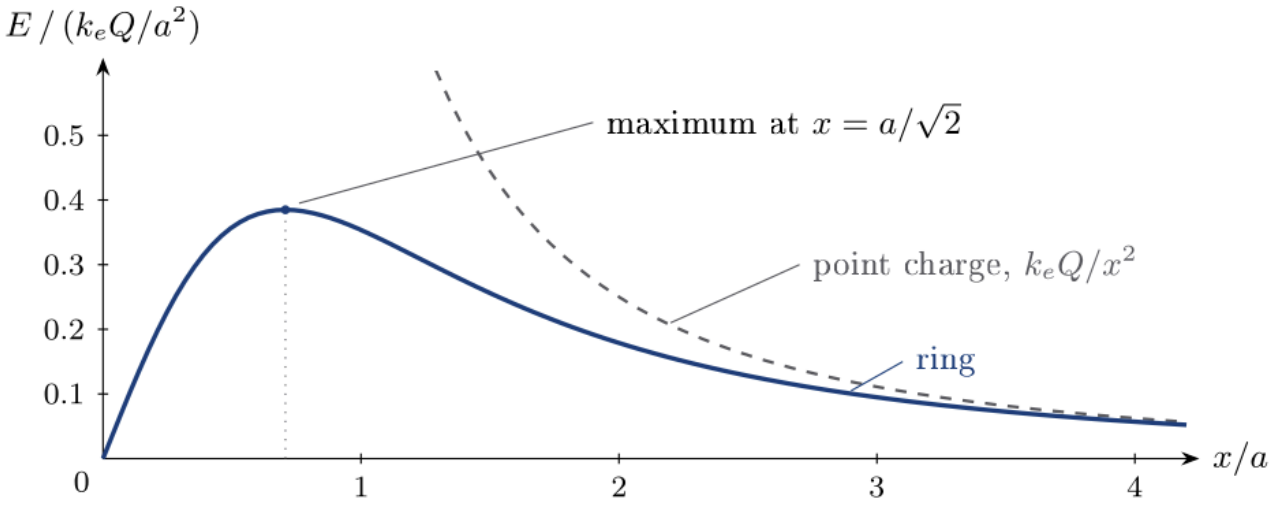


Figure 16.7. The field on the axis of a uniformly charged ring of radius a. It is zero at the center, peaks at $x = a/\sqrt{2}$, and approaches the point-charge field k_eQ/x^2 (dashed) at distances of a few radii.

•Figure description

A graph of the field on the axis of a charged ring, in units of k Q over a squared, against distance along the axis in units of the ring radius a, from 0 to 4. The curve starts at zero at the center of the ring, rises steeply to a maximum of about 0.385 at x equals a over root 2, about 0.71, and then falls away gradually. A dashed curve, one over x squared for a point charge, lies far above the ring's curve near the ring but merges with it beyond about x equals 3 a.

directed along the axis, away from the ring if Q is positive. Besides being useful in its own right, the ring is a building block: a disk is nothing but a set of concentric rings, and Section 16.3 uses equation (16.11) that way. Its shape shows several features that will recur. Figure 16.7 plots it, in units of k_eQ/a^2, against the distance x/a along the axis.

Three features stand out. At the center of the ring the field is zero, because the contributions from opposite sides cancel exactly. Far away, the field approaches that of a point charge. And in between there is a maximum. The field grows at first because moving away from the center stops the cancellation, and it eventually shrinks because everything is getting farther away.

Example 16.6: Where is a ring's field strongest?

Show that the field on the axis of a uniform ring of radius a is greatest at $x = a/\sqrt{2}$, and find the maximum field.

Solution

Differentiate equation (16.11) with respect to x using the product rule, and set the result to zero:

$$\frac{dE}{dx} = k_eQ\left[\frac{1}{(x^2+a^2)^{3/2}} - \frac{3x^2}{(x^2+a^2)^{5/2}}\right] = k_eQ\,\frac{a^2-2x^2}{(x^2+a^2)^{5/2}} = 0. \tag{16.3.6}$$

The numerator vanishes when $x^2 = a^2/2$, that is, at $x = \boxed{a/\sqrt{2}}$. Substituting back, with $x^2 + a^2 = 3a^2/2$,

$$E_{\max} = \frac{k_e Q\,(a/\sqrt{2})}{(3a^2/2)^{3/2}} = \frac{2}{3\sqrt{3}}\,\frac{k_e Q}{a^2} \approx \boxed{0.385\,\frac{k_e Q}{a^2}}. \tag{16.3.7}$$

Note: Off the axis

Every result in this section is for points on an axis of symmetry, and that is no accident. Off the axis, the contributions from different parts of the ring no longer cancel so neatly, and the integrals cannot be done in terms of elementary functions. Physicists and engineers find such fields numerically. The field-line pictures in this chapter were produced exactly that way, by adding the fields of point charges at many points and tracing the direction of the result.

Checkpoint 16.4

Near the center of a positively charged ring, the field on the axis points away from the center on both sides. (a) A small positive charge is released on the axis very close to the center. What happens to it? (b) What happens to a small negative charge released at the same point? *(Chapter 15, Problem 15.25, followed this case to the end.)*

Answer

The Electric Field of a Charged Disk

A thin disk of radius R carries charge spread uniformly over one face with surface charge density σ. We want the field at a point P on the disk's axis, a distance x from its center.

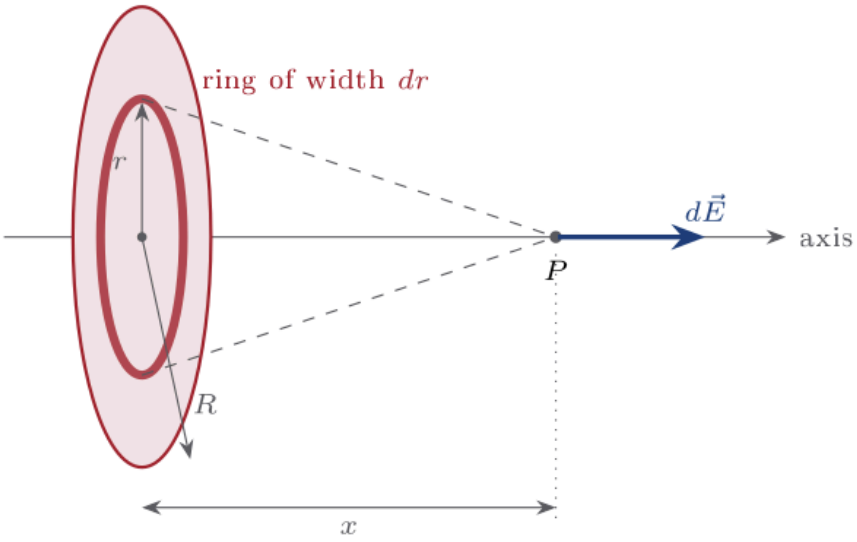


Figure 16.8. A uniformly charged disk treated as a set of concentric rings. The ring of radius r and width dr carries charge $dq = 2\pi\sigma r\ dr$ and, by equation (16.11), produces a field along the axis. All the rings' fields point the same way, so their magnitudes add.

•Figure description

A uniformly charged disk seen in perspective as a shaded ellipse, with its axis running horizontally to the right. A thin ring within the disk, of radius r and width d r, is highlighted. Point P lies on the axis a distance x from the disk's center, and the field d E of the highlighted ring at P points along the axis, away from the disk. The disk's full radius is labeled R.

Rather than integrating over tiny patches of area, which would need a double integral, we can use the ring result. Divide the disk into thin concentric rings (Figure 16.8). The ring of radius r and width dr has area $dA = 2\pi r\ dr$ and so carries charge

$$dq = \sigma\, dA = 2\pi\sigma r\, dr. \tag{16.3.8}$$

By equation (16.11), with a replaced by r and Q by dq, this ring produces a field along the axis of

$$dE = \frac{k_e x\, dq}{(x^2 + r^2)^{3/2}} = \frac{2\pi k_e \sigma x\, r\, dr}{(x^2 + r^2)^{3/2}}. \tag{16.3.9}$$

Every ring's field points along the axis, so the contributions simply add. Integrating from the center of the disk ($r = 0$) to its edge ($r = R$), with the substitution $u = x^2 + r^2$, $du = 2r\ dr$,

$$E = \pi k_e \sigma x \int_{x^2}^{x^2+R^2} u^{-3/2}\, du = \pi k_e \sigma x \left[-\frac{2}{\sqrt{u}}\right]_{x^2}^{x^2+R^2} = 2\pi k_e \sigma x \left(\frac{1}{x} - \frac{1}{\sqrt{x^2 + R^2}}\right). \tag{16.3.10}$$

Using $2\pi k_e = 1/(2\epsilon_0)$, the field on the axis of a uniformly charged disk is

Key Equation: Field on the axis of a charged disk

$$E = \frac{\sigma}{2\epsilon_0}\left(1 - \frac{x}{\sqrt{x^2 + R^2}}\right), \tag{16.12}$$

directed along the axis, away from the disk for positive σ.

Checks. Far away, $x \gg R$, we need the small difference between 1 and $x/\sqrt{x^2 + R^2} = (1 + R^2/x^2)^{-1/2} \approx 1 - R^2/2x^2$. Then

$$E \approx \frac{\sigma}{2\epsilon_0} \cdot \frac{R^2}{2x^2} = \frac{1}{4\pi\epsilon_0}\frac{\sigma\pi R^2}{x^2} = \frac{k_e Q}{x^2}, \tag{16.3.11}$$

since $\sigma\pi R^2$ is the disk's total charge Q. Once again, from far away the disk is a point charge. The near limit, $x \ll R$, leads to the infinite plane of Section 16.3.

Example 16.7: Field near a charged disk

A plastic disk of radius $10.0\ \text{cm}$ carries a uniform surface charge density of $5.0\ \mu\text{C/m}^2$ on one face. Find the field on its axis (a) $2.0\ \text{cm}$ and (b) $50\ \text{cm}$ from its center.

Solution

The prefactor is

$$\frac{\sigma}{2\epsilon_0} = \frac{5.0 \times 10^{-6}}{2(8.85 \times 10^{-12})} = 2.82 \times 10^5\ \text{N/C}. \tag{16.3.12}$$

(a) At $x = 0.020\ \text{m}$, $x/\sqrt{x^2 + R^2} = 0.020/0.102 = 0.196$, so

$$E = (2.82 \times 10^5)(1 - 0.196) = \boxed{2.3 \times 10^5\ \text{N/C}}. \tag{16.3.13}$$

(b) At $x = 0.50\ \text{m}$, $x/\sqrt{x^2 + R^2} = 0.50/0.5099 = 0.9806$, and

$$E = (2.82 \times 10^5)(1 - 0.9806) \approx \boxed{5.5 \times 10^3\ \text{N/C}}. \tag{16.3.14}$$

Here the answer is the small difference of two nearly equal numbers, so it pays to carry extra digits: rounding 0.9806 to 0.98 would have made the answer 3% too large. For comparison, the point-charge formula, with $Q = \sigma\pi R^2 = 1.57 \times 10^{-7}\ \text{C}$, gives $k_e Q/x^2 = 5.65 \times 10^3\ \text{N/C}$, already only about 3% high. Close to the disk, in part (a), the field is 80% of $\sigma/2\epsilon_0$, the value it approaches right at the surface.

The Electric Field of a Charged Plane

An **infinite plane of charge**, a flat sheet with uniform surface charge density σ extending forever in every direction, can be handled in two ways. Doing it both ways shows how much a little thought about *how* to slice up a distribution can simplify the integral.

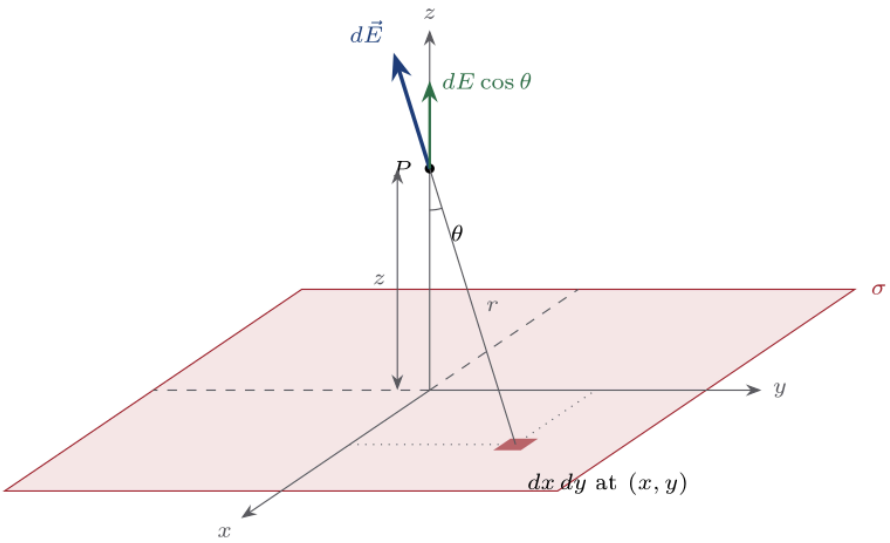


Figure 16.9. The direct route to the field of an infinite plane. The patch $dx\ dy$ at (x, y) is a distance r from P, and its field $d\mathbf{E}$ points along the line from the patch to P. The patch at $(-x, -y)$ cancels the part of $d\mathbf{E}$ parallel to the plane, so only the component $dE\cos\theta$ along the z-axis is added up.

•Figure description

An infinite plane of charge drawn in perspective as a shaded square lying flat, with x, y, and z axes meeting at its center. The z-axis points straight up. A small rectangular patch of the plane, of sides d x and d y, sits at the point x, y. Point P is on the z-axis a height z above the plane. A line of length r joins the patch to P and makes angle theta with the z-axis. At P, the field d E of the patch points along that line, away from the patch, and its vertical component d E cosine theta is drawn along the z-axis.

The direct route: a double integral. Put the plane in the xy-plane and find the field at a point P on the z-axis, a distance z above the plane (Figure 16.9). Slice the plane into small rectangles. The patch of sides dx and dy at the point (x, y) has area $dA = dx\ dy$, carries charge $dq = \sigma\ dx\ dy$, and is a distance

$$r = \sqrt{x^2 + y^2 + z^2} \tag{16.3.15}$$

from P. Its field $d\mathbf{E}$ points along the line from the patch to P. Now use symmetry. For every patch at (x, y) there is a matching patch at $(-x, -y)$, the same distance from P, whose field has the same z-component but opposite components parallel to the plane. Those parallel components cancel in pairs, and only the z-components survive. The line from the patch makes an angle θ with the z-axis, where $\cos\theta = z/r$, so

$$E = \int dE\cos\theta = \int k_e \frac{\sigma\, dx\, dy}{r^2} \cdot \frac{z}{r} = k_e \sigma z \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dy\, dx}{(x^2 + y^2 + z^2)^{3/2}}. \tag{16.3.16}$$

Do the inner integral over y first, holding x fixed. It sums the patches along one strip of the plane, parallel to the y-axis. With the abbreviation $b^2 = x^2 + z^2$, which is constant along the strip, a table of integrals (or differentiating the answer to check it) gives

$$\int_{-\infty}^{\infty} \frac{dy}{(y^2 + b^2)^{3/2}} = \left[\frac{y}{b^2\sqrt{y^2 + b^2}}\right]_{-\infty}^{\infty} = \frac{1}{b^2} - \left(-\frac{1}{b^2}\right) = \frac{2}{b^2} = \frac{2}{x^2 + z^2}. \tag{16.3.17}$$

That leaves an integral over x, which adds up the strips:

$$E = 2k_e\sigma z \int_{-\infty}^{\infty} \frac{dx}{x^2 + z^2} = 2k_e\sigma z \left[\frac{1}{z}\arctan\frac{x}{z}\right]_{-\infty}^{\infty} = 2k_e\sigma z \cdot \frac{1}{z}\left(\frac{\pi}{2} + \frac{\pi}{2}\right) = 2\pi k_e \sigma. \tag{16.3.18}$$

With $2\pi k_e = 1/(2\epsilon_0)$, this is $E = \sigma/2\epsilon_0$. The distance z has canceled completely.

The shortcut: let the disk grow. We already have the field of a disk. Let the disk of equation (16.12) grow without limit, $R \to \infty$. The fraction $x/\sqrt{x^2 + R^2}$ goes to zero, and what remains is $\sigma/2\epsilon_0$ again, in a single line.

Compare the two routes. The rectangular grid ignores the plane's symmetry about the axis through P, so we needed two integrals,

plus a symmetry argument to throw away the parallel components. It worked here only because the limits are infinite: for a finite disk the inner limits would become $\pm\sqrt{R^2 - x^2}$ and the integral would turn ugly. The rings of Section 16.3 match the symmetry. Every patch of a ring is the same distance from P, so the cancellation of the parallel components is already built in, and a single integral did the job. When a distribution has a symmetry, slice it into pieces that share that symmetry. Either way, and whatever we call the distance from the plane, the field of an infinite plane of charge is

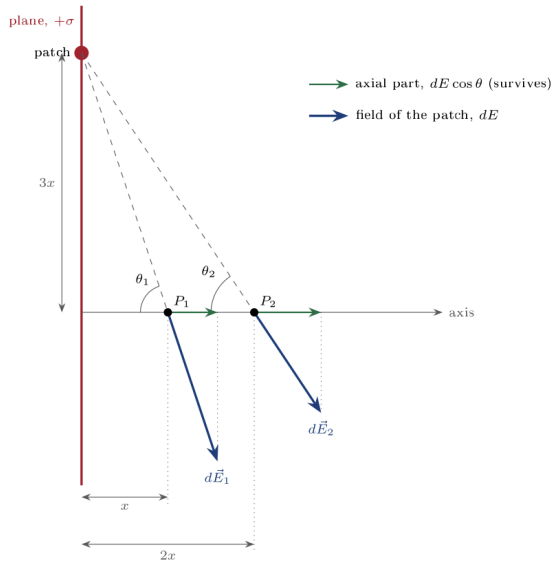


Figure 16.10. Why the field of an infinite plane does not weaken with distance. One patch of the plane, a distance $3x$ from the axis, is followed from P_1, a distance x from the plane, to P_2, at $2x$. At P_2 the patch is farther away and its field dE is weaker, but the line to it makes a smaller angle with the axis, so more of dE points along the axis. A matching patch the same distance below the axis (not drawn) cancels the rest. For this patch, the axial part (green) is about 35% *larger* at P_2 than at P_1. Patches close to the axis lose more than they gain, patches far out gain more than they lose, and over the whole infinite plane the two balance exactly.

•Figure description

A charged plane is drawn edge-on as a vertical line, with a horizontal axis perpendicular to it. A charge patch sits on the plane well above the axis. Two field points lie on the axis: P sub 1 close to the plane and P sub 2 twice as far away. Lines run from the patch to each point. At P sub 1 the line is steep, making a large angle theta sub 1 with the axis, and the patch's field there is a long arrow pointing mostly away from the axis, with only a short component along the axis. At P sub 2 the line is shallower, making a smaller angle theta sub 2, and the patch's field is a shorter arrow, but its component along the axis is longer than at P sub 1. A matching patch the same distance below the axis, not drawn, cancels the components perpendicular to the axis.

Key Equation: Field of an infinite plane of charge

$$E = \frac{\sigma}{2\epsilon_0}. \tag{16.13}$$

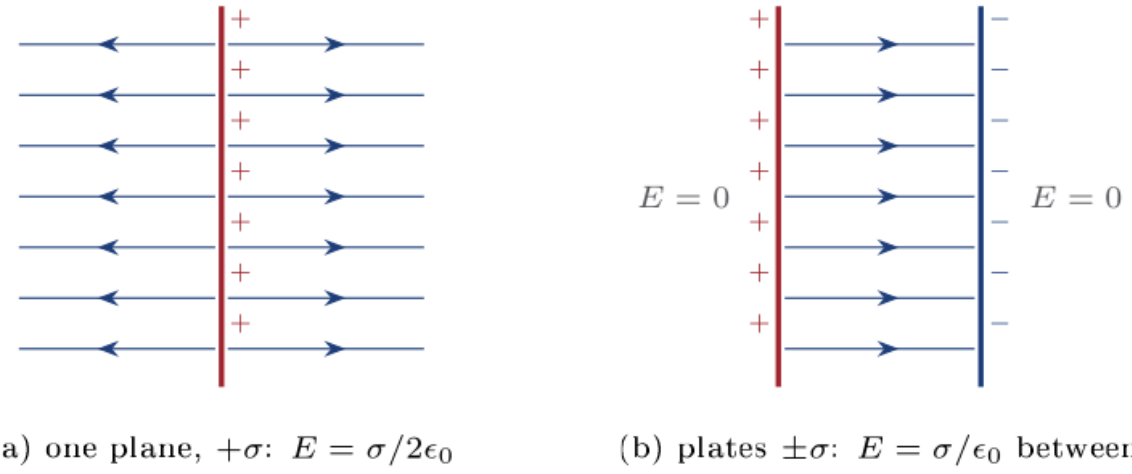


Figure 16.11. (a) A large plane of positive charge produces a uniform field $\sigma/2\epsilon_0$ pointing away from it on both sides. (b) Two parallel plates with charge densities $+\sigma$ and $-\sigma$: between them the two fields add to σ/ϵ_0, and outside they cancel.

•Figure description

Two panels. Panel a shows a single positively charged plane drawn edge-on as a vertical line, with evenly spaced parallel field lines pointing straight away from it on both sides. Panel b shows two parallel plates edge-on, the left plate positive and the right plate negative. Evenly spaced parallel field lines run from the positive plate to the negative plate in the region between them, and there are no field lines outside the plates.

The field is perpendicular to the plane, pointing away from it on both sides if σ is positive and toward it if σ is negative. The most striking thing about equation (16.13) is what it does not contain: the distance x. The field of an infinite plane is the same at every distance from it. Its field lines are parallel, evenly spaced straight lines that never spread out (Figure 16.11a), and a field that has the same magnitude and direction everywhere in a region is called a **uniform electric field**.

How can the field fail to weaken with distance? Figure 16.10 follows one patch of charge as the field point moves away from the plane. Two things change. The patch is farther away, so its field dE is weaker. But the line from the patch now meets the axis at a smaller angle θ, so a larger share of dE, the fraction $\cos\theta$, points along the axis and survives the cancellation. For patches near the foot of the perpendicular, θ is already small, and the first effect wins: their contribution shrinks. For patches far out across the plane, the second effect wins, and their contribution actually *grows* as you move away. For an infinite plane the gains and losses balance exactly.

No real sheet of charge is infinite, of course. Equation (16.13) is a good approximation for a finite sheet at points much closer to it than to its edges, $x \ll R$, as part (a) of the disk example showed.

Two parallel planes. The most useful arrangement is a pair of large parallel plates carrying equal and opposite charge densities, $+\sigma$ and $-\sigma$ (Figure 16.11b). Each plate alone produces a field of magnitude $\sigma/2\epsilon_0$. Between the plates, the positive plate's field points away from it and the negative plate's field points toward it, so both point from the positive plate to the negative one and they add. Outside the plates, on either side, the two fields point in opposite directions and cancel. Therefore

Key Equation: Field between oppositely charged parallel plates

$$E = \frac{\sigma}{\epsilon_0} \quad \text{between the plates,} \qquad E = 0 \quad \text{outside.} \tag{16.14}$$

The field between the plates is uniform and points from the positive plate to the negative plate. This arrangement is the standard way to produce a uniform field in the laboratory, and it is the heart of the capacitor (Chapter 19). Near the edges of real plates the field lines bulge outward, an effect called fringing, but for plates much larger than their separation it is small.

Example 16.8: A pair of charged plates

Two parallel square metal plates, each $10.0\ \text{cm}$ on a side, face each other a few millimeters apart. One carries $+5.0\ \text{nC}$ and the other $-5.0\ \text{nC}$, spread uniformly over their facing surfaces. Find the field between the plates.

Solution

The charge density on each plate has magnitude

$$\sigma = \frac{Q}{A} = \frac{5.0 \times 10^{-9}\,\mathrm{C}}{(0.100\,\mathrm{m})^2} = 5.0 \times 10^{-7}\,\mathrm{C/m^2}. \tag{16.3.19}$$

Since the plates are much larger than their separation, equation (16.14) applies:

$$E = \frac{\sigma}{\epsilon_0} = \frac{5.0 \times 10^{-7}}{8.85 \times 10^{-12}} = \boxed{5.6 \times 10^4\,\mathrm{N/C}}, \tag{16.3.20}$$

directed from the positive plate toward the negative plate, and the same everywhere between them. The separation does not enter, as long as it stays small compared with the size of the plates.

The Electric Field of a Charged Sphere

The last shape is the sphere. Because Coulomb's law has exactly the same inverse-square form as Newton's law of gravitation, the two shell theorems of Chapter 13 carry over to charge unchanged, with $k_e q$ in place of Gm:

1. A uniform spherical *shell* of charge produces, at every point outside it, the same field as a point charge at its center carrying the shell's total charge.
2. A uniform spherical shell of charge produces *zero* field everywhere inside it.

Newton proved these by an integration much like the ones in this chapter. Chapter 17 will prove them again in a few lines using Gauss's law. Here we use them to find the field of two important spheres.

A spherical shell, or a charged conducting sphere. If a charge Q is spread uniformly over a thin spherical shell of radius R, the shell theorems give the field at once:

$$E = \frac{k_e Q}{r^2} \quad (r > R), \qquad E = 0 \quad (r < R). \tag{16.3.21}$$

The same is true for a charged solid *conducting* sphere, because on a conductor the charge moves to the outer surface, as we will show in Chapter 17.

A uniformly charged solid sphere. Now suppose a charge Q is spread uniformly through the *volume* of a sphere of radius R, an insulating ball, say, with volume charge density $\rho = Q/(\frac{4}{3}\pi R^3)$. Outside the sphere, think of it as a set of nested shells. Each shell acts as a point charge at the center, so the whole sphere does too, and $E = k_e Q/r^2$.

Inside the sphere, at a distance $r < R$ from the center, divide the sphere into the part inside radius r and the shells outside it. The outer shells all enclose the point, so by the second shell theorem they contribute nothing. The inner part is a smaller uniform sphere, which acts as a point charge at the center carrying the charge it contains:

$$Q_{\mathrm{in}} = \rho \cdot \tfrac{4}{3}\pi r^3 = Q\,\frac{r^3}{R^3}. \tag{16.3.22}$$

The field at r is therefore $E = k_e Q_{\mathrm{in}}/r^2 = k_e Q r/R^3$. Putting both regions together:

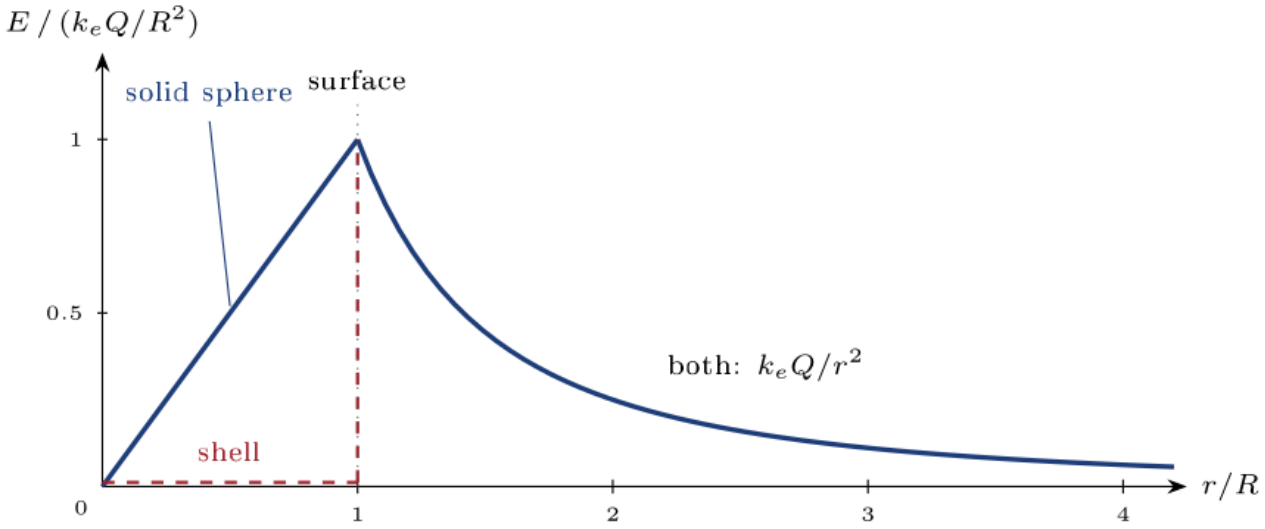


Figure 16.12. The field of a charged sphere of radius R. Solid: charge spread uniformly through the volume, with E growing linearly inside. Dashed: charge on a thin shell or a conductor's surface, with $E = 0$ inside. Outside, both are identical to the field of a point charge.

•Figure description

A graph of electric field against distance r from the center of a charged sphere of radius R, with r from 0 to 4 R. For a uniformly charged solid sphere, drawn as a solid curve, the field rises in a straight line from zero at the center to its maximum k Q over R squared at the surface, then falls off as one over r squared. For a thin shell or a conducting sphere, drawn dashed, the field is zero everywhere inside, jumps to the same maximum at the surface, and outside follows the same one over r squared curve.

Key Equation: Field of a uniformly charged solid sphere

$$E = \frac{k_eQ}{R^3}\,r \quad (r \le R), \qquad E = \frac{k_eQ}{r^2} \quad (r \ge R). \tag{16.15}$$

Inside, the field grows linearly from zero at the center. Outside, it falls off as $1/r^2$. The two expressions agree at the surface, where both give k_eQ/R^2, the largest field anywhere (Figure 16.12). The field is radial everywhere, outward for positive Q.

Example 16.9: Inside and outside a charged ball

An insulating ball of radius $5.0\ \text{cm}$ carries $+20\ \text{nC}$ spread uniformly through its volume. Find the field (a) at its center, (b) $2.5\ \text{cm}$ from its center, (c) at its surface, and (d) $10\ \text{cm}$ from its center.

Solution

(a) At the center, $r = 0$ and $\boxed{E = 0}$. Charge surrounds the center symmetrically, so every push is balanced.

(b) Inside, at $r = 0.025\ \text{m}$, equation (16.15) gives

$$E = \frac{k_eQ\,r}{R^3} = \frac{(8.99\times10^9)(20\times10^{-9})(0.025)}{(0.050)^3} = \boxed{3.6\times10^4\,\text{N/C}}. \tag{16.3.23}$$

(c) At the surface,

$$E = \frac{k_eQ}{R^2} = \frac{(8.99\times10^9)(20\times10^{-9})}{(0.050)^2} = \boxed{7.2\times10^4\,\text{N/C}}. \tag{16.3.24}$$

(d) Outside, at $r = 0.10\ \text{m}$,

$$E = \frac{k_eQ}{r^2} = \frac{(8.99\times10^9)(20\times10^{-9})}{(0.10)^2} = \boxed{1.8\times10^4\,\text{N/C}}. \tag{16.3.25}$$

Halfway to the surface from the center the field is half its surface value, and twice the radius away it is a quarter of it: linear inside, inverse-square outside. All four fields point radially outward.

Checkpoint 16.5

Rank the magnitudes of the fields at these four points, from largest to smallest: (a) $1\ \mathrm{cm}$ from an infinite plane with $\sigma = 1\ \mu\mathrm{C/m^2}$; (b) $1\ \mathrm{m}$ from the same plane; (c) midway between two large plates carrying $\pm 1\ \mu\mathrm{C/m^2}$; (d) $1\ \mathrm{cm}$ outside those two plates. Explain each answer without calculating.

Answer

Answers to Checkpoints

Checkpoint 16.4. (a) The field points away from the center on both sides, so a positive charge placed just off center is pushed *farther* from the center; it accelerates away along the axis. The center is an unstable equilibrium for it. (b) A negative charge feels a force opposite to the field, back toward the center. It is pulled through the center, overshoots, is pulled back, and oscillates about the center along the axis; for small displacements the motion is simple harmonic.

Checkpoint 16.5. From largest to smallest: (c), then (a) and (b) tied, then (d). The field of an infinite plane, $\sigma/2\epsilon_0$, does not depend on distance, so (a) and (b) are equal. Between two oppositely charged plates the two plates' fields point the same way and add, giving σ/ϵ_0: twice (a). Outside the pair the two fields point in opposite directions and cancel, so (d) is essentially zero.

This page titled 16.3: The Electric Field of Continuous Charge Distributions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.4: A Point Charge in an Electric Field

We now turn the question around. Instead of finding the field produced by charges, suppose the field is known and ask what it does to a charged particle placed in it. By equation (16.2) the force on a particle of charge q is $q\mathbf{E}$. If this is the only force that matters, Newton's second law gives the particle's acceleration:

Key Equation: Acceleration of a charge in an electric field

$$\mathbf{a} = \frac{q\mathbf{E}}{m}. \tag{16.16}$$

The acceleration depends on the particle only through its **charge-to-mass ratio** q/m. For electrons, $e/m_e = 1.76 \times 10^{11}\ \mathrm{C/kg}$, an enormous number, and even modest fields give them tremendous accelerations. For the same reason, the gravitational force on an electron or proton is almost always negligible next to the electric force. It matters only for macroscopic charged objects such as dust grains and oil drops.

In a *uniform* field, $\mathbf{a}$ is constant, and everything you learned about motion with constant acceleration in Chapters 2 and 4 applies directly. A charge released from rest accelerates in a straight line along (or against) the field. A charge that enters the field moving sideways follows a parabola, exactly like a projectile, with $q\mathbf{E}/m$ taking the place of $\mathbf{g}$.

Example 16.10: Accelerating an electron

An electron is released from rest at the negative plate of a pair of parallel plates $1.50\ \mathrm{cm}$ apart, between which there is a uniform field of $2.00 \times 10^4\ \mathrm{N/C}$. (a) What is its acceleration? (b) How fast is it moving when it reaches the positive plate, and (c) how long does the trip take?

Solution

(a) The field points from the positive plate to the negative plate. The electron is negative, so the force on it points the other way, toward the positive plate, as expected. From equation (16.16),

$$a = \frac{eE}{m_e} = \frac{(1.602 \times 10^{-19})(2.00 \times 10^4)}{9.11 \times 10^{-31}} = \boxed{3.52 \times 10^{15}\,\mathrm{m/s^2}}. \tag{16.4.1}$$

(b) With constant acceleration from rest over $\Delta x = 0.0150\ \mathrm{m}$, $v^2 = 2a\ \Delta x$:

$$v = \sqrt{2(3.52 \times 10^{15})(0.0150)} = \boxed{1.03 \times 10^7\,\mathrm{m/s}}, \tag{16.4.2}$$

about 3% of the speed of light.

(c) From $v = at$,

$$t = \frac{v}{a} = \frac{1.03 \times 10^7}{3.52 \times 10^{15}} = \boxed{2.92 \times 10^{-9}\,\mathrm{s}}. \tag{16.4.3}$$

The whole journey takes three nanoseconds. For comparison, the gravitational force on the electron, $m_e g \approx 9 \times 10^{-30}\ \mathrm{N}$, is some 10^{14} times smaller than the electric force $eE = 3.2 \times 10^{-15}\ \mathrm{N}$, and ignoring it was more than justified.

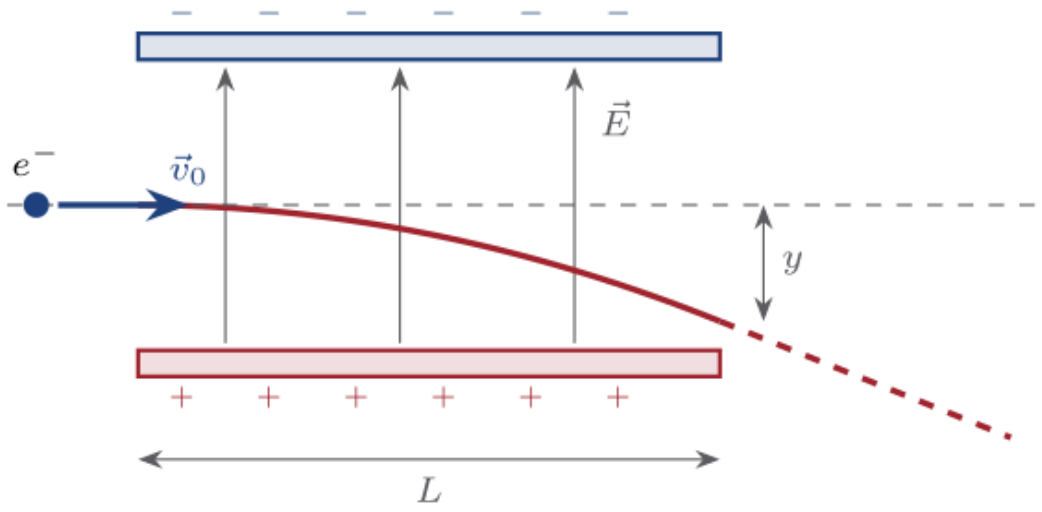


Figure 16.13. An electron entering a uniform field at right angles. The force $-e\mathbf{E}$ is constant and perpendicular to $\mathbf{v}_0$, so between the plates the electron follows a parabola, exactly like a horizontally launched projectile. Beyond the plates there is no field and it moves in a straight line.

•Figure description

Two horizontal parallel plates, the upper one negative and the lower one positive, so the field E between them points upward. An electron enters from the left midway between the plates with horizontal velocity v sub zero. Because the force on the negative electron is opposite to E, its path curves downward in a parabola while between the plates, and after leaving the plates it continues in a straight line angled downward. The plate length L and the vertical deflection y are marked.

Example 16.11: Deflecting an electron beam

An electron enters the region between two horizontal plates moving horizontally at $v_0 = 3.00 \times 10^7\ \text{m/s}$, midway between them (Figure 16.13). The plates are $L = 5.00\ \text{cm}$ long, and between them the field is uniform, $2.00 \times 10^4\ \text{N/C}$, pointing upward. (a) How far has the electron been deflected when it leaves the plates? (b) At what angle is it then moving?

Solution

Take $+x$ along $\mathbf{v}_0$ and $+y$ upward. The force on the electron points down, and from the previous example its acceleration is $a_y = -3.52 \times 10^{15}\ \text{m/s}^2$. Horizontally there is no force, so the electron keeps its speed v_0 and spends a time

$$t = \frac{L}{v_0} = \frac{0.0500\,\text{m}}{3.00 \times 10^7\,\text{m/s}} = 1.67 \times 10^{-9}\,\text{s} \tag{16.4.4}$$

between the plates. This is exactly a horizontally launched projectile.

(a) The vertical displacement is

$$y = \tfrac{1}{2}a_y t^2 = \tfrac{1}{2}(-3.52 \times 10^{15})(1.67 \times 10^{-9})^2 = \boxed{-4.9 \times 10^{-3}\,\text{m}}, \tag{16.4.5}$$

about $4.9\ \text{mm}$ downward.

(b) The vertical velocity on leaving is $v_y = a_y t = -5.86 \times 10^6\ \text{m/s}$, so the electron moves at an angle

$$\theta = \arctan\frac{|v_y|}{v_0} = \arctan\frac{5.86 \times 10^6}{3.00 \times 10^7} = \boxed{11.1^\circ} \tag{16.4.6}$$

below the horizontal. By adjusting the field, one can steer the beam. This is how the older cathode-ray tubes in oscilloscopes and televisions aimed their electron beams, and inkjet printers steer charged droplets of ink in the same way.

Example 16.12: Millikan's oil drop

In Millikan's experiment (Chapter 15), a tiny oil drop of mass $3.60 \times 10^{-15}\ \text{kg}$ is held motionless between two horizontal plates by a uniform vertical field of $7.34 \times 10^{4}\ \text{N/C}$. The drop carries excess electrons. (a) In which direction must the field point? (b) How many excess electrons does the drop carry?

Solution

(a) The electric force must balance gravity, so it must point up. The drop is negative, and the force on a negative charge is opposite to $\mathbf{E}$, so the field must point $\boxed{\text{downward}}$.

(b) Balancing the forces, $|q|E = mg$:

$$|q| = \frac{mg}{E} = \frac{(3.60 \times 10^{-15})(9.80)}{7.34 \times 10^{4}} = 4.81 \times 10^{-19}\,\text{C}, \tag{16.4.7}$$

and dividing by e,

$$n = \frac{4.81 \times 10^{-19}}{1.602 \times 10^{-19}} = 3.00 \implies \boxed{3\ \text{electrons}}. \tag{16.4.8}$$

Millikan measured many drops this way (in practice, he found each drop's mass from how fast it fell through the air) and found that every charge was a whole-number multiple of the same value, e.

Checkpoint 16.6

A proton and an electron are released from rest in the same uniform electric field. (a) Compare the magnitudes and directions of the forces on them. (b) Compare their accelerations. (c) After each has traveled $1\ \text{cm}$, which is moving faster, and by roughly what factor? *(Hint: $v^2 = 2a\ \Delta x$, and $m_p/m_e \approx 1840$.)*

Answer

Answers to Checkpoints

Checkpoint 16.6. (a) The forces have the same magnitude, eE, since the two charges have the same size; they point in opposite directions, the proton's along $\mathbf{E}$ and the electron's against it. (b) With equal forces, $a = F/m$, so the electron's acceleration is larger by the mass ratio, about 1840 times. (c) From $v^2 = 2a\ \Delta x$ with the same Δx, $v \propto \sqrt{a}$. The electron is faster by about $\sqrt{1840} \approx 43$.

This page titled 16.4: A Point Charge in an Electric Field is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.5: A Dipole in an Electric Field

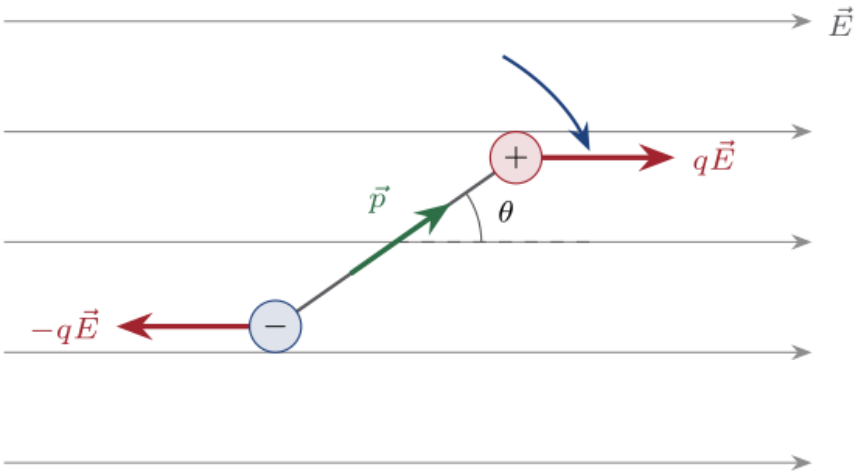


Figure 16.14. A dipole in a uniform field. The forces on the two charges are equal and opposite, so the net force is zero, but they form a couple whose torque $\tau = pE\sin\theta$ turns $\mathbf{p}$ toward the direction of $\mathbf{E}$.

•Figure description

An electric dipole in a uniform field E that points to the right. The dipole's axis is tilted at angle theta above the field direction, with the positive charge at the upper right and the negative charge at the lower left, and the dipole moment p points from the negative toward the positive charge. The force on the positive charge, q E, points to the right; the force on the negative charge, minus q E, points to the left. The two forces form a couple that rotates the dipole clockwise, toward alignment with the field.

Finally, place an electric dipole, with charges $\pm q$ a distance d apart, in a *uniform* external field $\mathbf{E}$, with its dipole moment $\mathbf{p}$ at an angle θ to the field (Figure 16.14). The positive charge feels a force $q\mathbf{E}$ along the field and the negative charge feels $-q\mathbf{E}$, equal and opposite. The net force on the dipole is zero, so its center of mass does not accelerate. The two forces do not act along the same line, however, so they exert a torque. About the center of the dipole, each force has a moment arm $(d/2)\sin\theta$, and both turn the dipole the same way, toward alignment with the field. The magnitude of the net torque is

$$\tau = 2\,(qE)\left(\frac{d}{2}\sin\theta\right) = qdE\sin\theta = pE\sin\theta. \tag{16.5.1}$$

In the vector language of Chapter 10 this is a cross product:

Key Equation: Torque on an electric dipole

$$\tau = \mathbf{p}\times\mathbf{E}. \tag{16.17}$$

The torque is zero when $\mathbf{p}$ is parallel to $\mathbf{E}$ ($\theta = 0$) or antiparallel to it ($\theta = 180^\circ$), and largest, pE, when the dipole is perpendicular to the field. It always tends to rotate $\mathbf{p}$ toward the direction of $\mathbf{E}$.

Potential energy. As the field turns the dipole, it does work on it, so a dipole in a field has potential energy that depends on its orientation. From Chapter 11, a torque does work $dW = \tau\ d\theta$ as the dipole rotates, and the change in potential energy is the negative of the work done by the field. Since the field's torque acts to decrease θ, it is $-pE\sin\theta$ in the direction of increasing θ, and

$$\Delta U = -W = -\int_{\theta_1}^{\theta_2}(-pE\sin\theta)\,d\theta = -pE\cos\theta_2 + pE\cos\theta_1. \tag{16.5.2}$$

Only changes in U matter, so we are free to choose where $U = 0$. The convenient choice is $U = 0$ at $\theta = 90^\circ$, which gives

Key Equation: Potential energy of an electric dipole

$$U = -pE\cos\theta = -\mathbf{p}\cdot\mathbf{E}. \tag{16.18}$$

The potential energy is lowest, $-pE$, when the dipole is aligned with the field. This is a stable equilibrium: tilt the dipole slightly and the torque pulls it back, so it oscillates about $\theta = 0$ like a pendulum (Problem 16.29). The energy is highest, $+pE$, when the dipole points against the field. That is also an equilibrium, since the torque is zero there, but it is unstable: the slightest tilt, and the dipole swings all the way around. Turning a dipole from the stable to the unstable orientation takes work $2pE$.

Example 16.13: A water molecule in a strong field

A water molecule, $p = 6.2 \times 10^{-30}\ \mathrm{C\cdot m}$, sits in a field of $2.0 \times 10^{5}\ \mathrm{N/C}$. (a) What is the largest torque the field can exert on it? (b) How much work does it take to turn the molecule from alignment with the field to the opposite orientation? (c) Compare your answer to (b) with the typical thermal energy of a molecule at room temperature, $k_B T \approx 4.1 \times 10^{-21}\ \mathrm{J}$ (Chapter 36).

Solution

(a) The torque is greatest at $\theta = 90^\circ$:

$$\tau_{\max} = pE = (6.2 \times 10^{-30})(2.0 \times 10^{5}) = \boxed{1.2 \times 10^{-24}\,\mathrm{N\cdot m}}. \tag{16.5.3}$$

(b) From equation (16.18), the work is the difference in potential energy between $\theta = 180^\circ$ and $\theta = 0$:

$$W = U(180^\circ) - U(0) = pE - (-pE) = 2pE = \boxed{2.5 \times 10^{-24}\,\mathrm{J}}. \tag{16.5.4}$$

(c) This is more than a thousand times smaller than $k_B T$. The random jostling of molecules at room temperature easily overwhelms the aligning torque, so even in a field this strong, the water molecules are only very slightly more likely to point along the field than against it. It takes fields of order $10^{9}\ \mathrm{N/C}$, like those very close to an ion, to line water molecules up completely.

Note: Microwave ovens

A microwave oven fills its interior with an electric field that reverses direction about five billion times a second (twice per cycle at $2.45\ \mathrm{GHz}$). The torque of equation (16.17) tries to turn each water molecule in the food first one way and then the other. As the molecules twist, they jostle their neighbors, and the energy the field puts in ends up as random thermal motion. The food gets hot. Materials without polar molecules, such as glass, most plastics, and dry paper, are warmed very little.

Dipoles in a nonuniform field. In a uniform field the forces on a dipole's two charges cancel. In a nonuniform field they do not, because one end sits in a stronger field than the other, and a net force remains. A dipole aligned with a nonuniform field is pulled toward the region where the field is *stronger*. This explains a puzzle from Chapter 15. The field of a charged comb is strongest near the comb. A scrap of paper held nearby is polarized, turning into a set of small dipoles aligned with the field (Figure 15.4), and those dipoles are drawn toward the strong-field region, which is to say toward the comb, whatever the sign of its charge.

Checkpoint 16.7

An electric dipole is placed in a uniform field with $\mathbf{p}$ perpendicular to $\mathbf{E}$ and released. (a) What is the net force on it? (b) Describe its subsequent motion if there is no friction. (c) Through what angle does it turn before its kinetic energy is greatest?

Answer

Answers to Checkpoints

Checkpoint 16.7. (a) Zero. In a uniform field the forces $+q\mathbf{E}$ and $-q\mathbf{E}$ on the two ends are equal and opposite. (b) There is a torque, which rotates $\mathbf{p}$ toward $\mathbf{E}$. The dipole swings through the aligned orientation, overshoots to 90° on the other side, stops, and swings back, oscillating like a pendulum about alignment with $\mathbf{E}$. (c) 90°. Its potential energy, $-pE\cos\theta$, is lowest when $\mathbf{p}$ is

aligned with $\mathbf{E}$, so its kinetic energy is greatest there, after it has turned through 90°.

This page titled 16.5: A Dipole in an Electric Field is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.S: Electric Fields (Summary)

Summary

- The **electric field** at a point is the force per unit charge on a small positive test charge placed there, $\mathbf{E} = \mathbf{F}/q_0$, measured in $\mathrm{N/C}$. A charge q at that point feels $\mathbf{F} = q\mathbf{E}$. Charges create fields, and fields exert forces on charges.
- **Field lines** are tangent to $\mathbf{E}$, begin on positive charges and end on negative ones, and crowd together where the field is strong. They never cross.
- A point charge has field $\mathbf{E} = k_e(q/r^2)\ \hat{r}$, and fields of several charges add as vectors. An **electric dipole** of moment $p = qd$ (pointing from $-q$ to $+q$) produces a field that falls off as $1/r^3$ far away: $2k_ep/r^3$ on its axis and k_ep/r^3 on its perpendicular bisector.
- For a continuous distribution, $\mathbf{E} = k_e\int(dq/r^2)\ \hat{r}$, with symmetry used to cancel components. On axes of symmetry:
  - a line of charge: $E = k_eQ/\left(y\sqrt{y^2 + L^2/4}\right)$, becoming $\lambda/2\pi\epsilon_0 r$ for an infinite line;
  - a ring: $E = k_eQx/(x^2 + a^2)^{3/2}$, maximum at $x = a/\sqrt{2}$;
  - a disk: $E = (\sigma/2\epsilon_0)\left(1 - x/\sqrt{x^2 + R^2}\right)$;
  - an infinite plane: $E = \sigma/2\epsilon_0$, uniform; two oppositely charged plates: σ/ϵ_0 between them and zero outside;
  - a uniformly charged solid sphere: $E = k_eQr/R^3$ inside and k_eQ/r^2 outside. A shell or conducting sphere has $E = 0$ inside.

  Every one of these reduces to the point-charge field far away.

- A charge in a field accelerates at $\mathbf{a} = q\mathbf{E}/m$. In a uniform field the acceleration is constant, and the motion is that of a projectile with $q\mathbf{E}/m$ in place of $\mathbf{g}$.
- A dipole in a uniform field feels no net force but a torque $\tau = \mathbf{p} \times \mathbf{E}$, and has potential energy $U = -\mathbf{p} \cdot \mathbf{E}$, lowest when aligned with the field. In a nonuniform field an aligned dipole is pulled toward the stronger field.

Key Equations

Electric field, and force on a charge in a field

$$\mathbf{E} = \frac{\mathbf{F}}{q_0}, \qquad \mathbf{F} = q\mathbf{E} \tag{16.S.1}$$

Field of a point charge, and superposition

$$\mathbf{E} = k_e\,\frac{q}{r^2}\,\hat{r}, \qquad \mathbf{E} = \sum_i k_e\,\frac{q_i}{r_i^2}\,\hat{r}_i \tag{16.S.2}$$

Dipole moment, and dipole field at large distance

$$p = qd, \qquad E_{\parallel} \approx \frac{2k_ep}{r^3}, \qquad E_{\perp} \approx \frac{k_ep}{r^3} \tag{16.S.3}$$

Field of a continuous charge distribution

$$\mathbf{E} = k_e\int\frac{dq}{r^2}\,\hat{r} \tag{16.S.4}$$

Infinite line of charge

$$E = \frac{\lambda}{2\pi\epsilon_0 r} \tag{16.S.5}$$

Axis of a charged ring

$$E = \frac{k_e Q\, x}{\left(x^2 + a^2\right)^{3/2}} \tag{16.S.6}$$

Axis of a charged disk

$$E = \frac{\sigma}{2\epsilon_0}\left(1 - \frac{x}{\sqrt{x^2 + R^2}}\right) \tag{16.S.7}$$

Infinite plane, and oppositely charged parallel plates

$$E = \frac{\sigma}{2\epsilon_0}, \qquad E = \frac{\sigma}{\epsilon_0} \text{ (between the plates)} \tag{16.S.8}$$

Uniformly charged solid sphere

$$E = \frac{k_e Q}{R^3}\, r \quad (r \le R), \qquad E = \frac{k_e Q}{r^2} \quad (r \ge R) \tag{16.S.9}$$

Acceleration of a charge in a field

$$\mathbf{a} = \frac{q\mathbf{E}}{m} \tag{16.S.10}$$

Torque on, and energy of, a dipole in a field

$$\tau = \mathbf{p} \times \mathbf{E}, \qquad U = -\mathbf{p} \cdot \mathbf{E} \tag{16.S.11}$$

This page titled 16.S: Electric Fields (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

16.E: Electric Fields (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $e = 1.602 \times 10^{-19}$ C, $k_e = 8.99 \times 10^{9}$ N·m^2/C^2, $\epsilon_0 = 8.85 \times 10^{-12}$ C^2/(N·m^2), $m_e = 9.11 \times 10^{-31}$ kg, and $m_p = 1.673 \times 10^{-27}$ kg.

Section 16.1 — The Electric Field

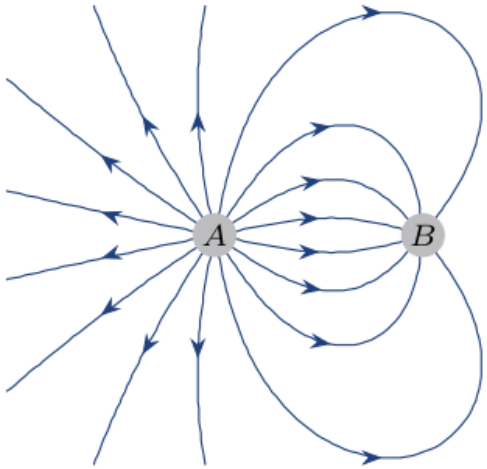
16.1 ★ A charge of -4.0 μC placed at a certain point feels a force of 0.012 N pointing straight up. Find the magnitude and direction of the electric field at that point.

16.2 ★ *[Real-world: atmospheric physics.]* Near the ground on a fair day, the Earth's electric field is about 150 N/C, pointing straight down.

a. Find the electric force on an electron in this field, and its direction.
b. Compare that force with the electron's weight.
c. What charge would a 1.0 g object need for the electric force to balance its weight?

16.3 ★★ *[Conceptual.]* Explain why the test charge used to measure an electric field must be small. Describe a situation in which a large test charge would give a wrong answer for the field that existed before it was brought in.

16.4 ★★ *[Field-line reading.]* The figure shows the computed field lines of two point charges, q_A on the left and q_B on the right.



•Figure description
Computed field lines of two point charges. The left charge, labeled A, has sixteen lines leaving it in all directions. The right charge, labeled B, has eight lines arriving at it, all of them coming from A; the two outermost of these swing around to reach B from its far side. The other eight lines from A curve outward and leave the drawing.

a. What are the signs of q_A and q_B?
b. What is the ratio $|q_A|/|q_B|$? How can you tell?
c. Far from both charges, what does the field look like? What single charge would produce nearly the same distant field?

Section 16.2 — The Electric Field of Point Charges

16.5 ★ Find the electric field (a) 1.0 m and (b) 1.0 mm from a point charge of $+1.0$ nC.

16.6 ★ *[Real-world: atomic physics.]* Find the electric field of the proton at the electron's orbit in hydrogen, $r = 5.29 \times 10^{-11}$ m, and check the value in Table 16.1.

16.7 ★★ A charge of $+3.0\ \text{nC}$ sits at the origin and a charge of $-5.0\ \text{nC}$ sits at $x = 0.40\ \text{m}$.

a. Find the electric field at $x = 0.20\ \text{m}$.
b. Find the electric field at $x = 0.60\ \text{m}$.
c. Is there a point on the x-axis where the field is zero? If so, in which region, and where is it?

16.8 ★★ Three charges of $+2.0\ \text{nC}$ each sit at three corners of a square $0.10\ \text{m}$ on a side. Find the magnitude and direction of the electric field at the fourth corner.

16.9 ★★ An electric dipole consists of charges $\pm 2.0\ \text{nC}$ separated by $1.0\ \text{mm}$.

a. Find its dipole moment.
b. Find the field $5.0\ \text{cm}$ from its center, on its axis and on its perpendicular bisector.
c. Use the exact expressions to find how large the error in the approximate results of (b) is.

16.10 ★★★ Show directly, from $E_{\parallel} = 2k_e q dy/(y^2 - d^2/4)^2$, that the axial field of a dipole approaches $2k_e p/y^3$ for $y \gg d$. Then find the next correction: show that $E_{\parallel} \approx (2k_e p/y^3)(1 + d^2/2y^2)$.

Section 16.3 — The Electric Field of Continuous Charge Distributions

16.11 ★ A long straight wire carries a uniform charge of $3.0\ \text{nC}$ per meter. Find the electric field $2.0\ \text{cm}$ from the wire.

16.12 ★ A large, flat, horizontal sheet carries a uniform surface charge of $+2.5\ \mu\text{C/m}^2$. Find the electric field (a) $1.0\ \text{cm}$ above the sheet and (b) $1.0\ \text{cm}$ below it. Give directions.

16.13 ★★ A rod $12\ \text{cm}$ long carries $+8.0\ \text{nC}$ spread uniformly. Find the electric field at a point $4.0\ \text{cm}$ from the rod's center on its perpendicular bisector. Compare your answer with the infinite-line approximation.

16.14 ★★ A thin ring of radius $3.0\ \text{cm}$ carries $+12\ \text{nC}$.

a. Find the field on its axis $4.0\ \text{cm}$ from its center.
b. Find the maximum field anywhere on the axis, and where it occurs.

16.15 ★★ A disk of radius $8.0\ \text{cm}$ carries a uniform surface charge density of $3.0\ \mu\text{C/m}^2$. Find (a) the field on its axis $1.0\ \text{cm}$ from its center, and (b) the distance from the center, along the axis, at which the field has fallen to half its value at the surface, and the field there.

16.16 ★★ *[Real-world: laboratory equipment.]* The sphere of a Van de Graaff generator has a radius of $20\ \text{cm}$. Air breaks down, and sparks fly, when the field reaches $3.0 \times 10^6\ \text{N/C}$. What is the largest charge the sphere can hold?

16.17 ★★ *[Graph-reading.]* A student measures the electric field at several distances from the center of a charged ball of radius R and plots E against r. Inside the ball the plot is a straight line through the origin, reaching $9.0 \times 10^4\ \text{N/C}$ at $r = R = 4.0\ \text{cm}$. Beyond that, the field falls to $2.25 \times 10^4\ \text{N/C}$ at $r = 8.0\ \text{cm}$.

a. Is the charge on the ball spread through its volume, or on its surface? How can you tell?
b. Find the total charge on the ball.
c. What field would the student measure at $r = 1.0\ \text{cm}$?

16.18 ★★★ A thin rod bent into a quarter circle of radius R carries a total charge Q spread uniformly. Find the magnitude and direction of the electric field at the center of the circle. *(Hint: take the arc from $\phi = 0$ to $\phi = 90^\circ$, and integrate each component separately.)*

16.19 ★★★ A semi-infinite line of charge, with uniform density λ, runs along the x-axis from $x = 0$ to $x = \infty$. Show that at the point $(0,\ y)$ on the y-axis the field has equal x- and y-components, each of magnitude $k_e\lambda/y$, so that it points at 45° to the line.

16.20 ★★★ Two infinite planes of charge cross at right angles, both carrying the same density $+\sigma$. Describe the electric field in each of the four regions they divide space into, giving its magnitude and direction.

Section 16.4 — A Point Charge in an Electric Field

16.21 ★ A proton is released from rest in a uniform field of $5.0 \times 10^{3}\ \mathrm{N/C}$.

a. Find its acceleration.
b. How fast is it moving after $2.0\ \mu\mathrm{s}$, and how far has it gone?

16.22 ★★ An electron moving at $4.0 \times 10^{6}\ \mathrm{m/s}$ enters a region where the field is uniform and points in the same direction as the electron's velocity, with magnitude $1.5 \times 10^{3}\ \mathrm{N/C}$.

a. Does the electron speed up or slow down?
b. How far does it travel before stopping momentarily?
c. How long does that take?

16.23 ★★ *[Real-world: printing.]* In an inkjet printer, an ink drop of mass $1.5 \times 10^{-10}\ \mathrm{kg}$ and charge $-2.0 \times 10^{-13}\ \mathrm{C}$ enters the gap between two deflecting plates moving horizontally at $20\ \mathrm{m/s}$. The plates are $1.6\ \mathrm{cm}$ long, and the field between them is $8.0 \times 10^{5}\ \mathrm{N/C}$, pointing upward.

a. Find the drop's vertical deflection on leaving the plates. (Ignore gravity, and then check that this was reasonable.)
b. How would the deflection change for a drop with twice the charge?

16.24 ★★ In a repeat of Millikan's experiment, a drop of mass $2.4 \times 10^{-15}\ \mathrm{kg}$ is held stationary by a field of $2.94 \times 10^{4}\ \mathrm{N/C}$ pointing upward. What is the sign of its charge, and how many elementary charges does it carry?

16.25 ★★★ An electron is launched at $5.0 \times 10^{6}\ \mathrm{m/s}$ at 45° above the horizontal into a region where a uniform field of $2.0 \times 10^{3}\ \mathrm{N/C}$ points straight up.

a. Find the electron's acceleration.
b. How long does it take to return to its launch height, and how far away does it land?
c. Explain why this is the same problem as projectile motion, and identify what plays the role of g.

Section 16.5 — A Dipole in an Electric Field

16.26 ★ A dipole with moment $4.0 \times 10^{-12}\ \mathrm{C \cdot m}$ sits in a uniform field of $5.0 \times 10^{4}\ \mathrm{N/C}$, with its moment at 30° to the field. Find (a) the torque on it and (b) its potential energy.

16.27 ★★ *[Conceptual.]* A dipole in a uniform field feels a torque but no net force. In a nonuniform field it feels both. Explain, with a sketch, why an aligned dipole is pulled toward the region of stronger field, and use this to explain why a charged balloon attracts small bits of paper.

16.28 ★★ How much work does it take to rotate a dipole of moment $3.0 \times 10^{-9}\ \mathrm{C \cdot m}$ in a uniform field of $4.0 \times 10^{5}\ \mathrm{N/C}$ (a) from $\theta = 0$ to $\theta = 90^{\circ}$, and (b) from $\theta = 0$ to $\theta = 180^{\circ}$?

16.29 ★★★ A dipole of moment p and moment of inertia I about its center sits in a uniform field E. It is rotated by a small angle θ from alignment and released.

a. Show that, for small θ, the torque is approximately $-pE\theta$.
b. Compare with the torsion pendulum of Chapter 28, and show that the dipole undergoes simple harmonic motion with angular frequency $\omega = \sqrt{pE/I}$.
c. Evaluate the period for a dipole made of two small spheres, each of mass $2.0\ \mathrm{g}$ and charge $\pm 10\ \mathrm{nC}$, on a light rod $4.0\ \mathrm{cm}$ long, in a field of $1.0 \times 10^{5}\ \mathrm{N/C}$.

16.30 ★★★ *[Real-world: atmospheric physics.]* The Earth carries a net negative charge, which produces the fair-weather field of about $150\ \mathrm{N/C}$ pointing down at the surface.

a. Treating the Earth as a conducting sphere of radius $6.37 \times 10^{6}\ \mathrm{m}$, find the charge it carries.
b. Find the corresponding surface charge density, and check that $E = \sigma/\epsilon_0$ just above the surface. *(Chapter 17 shows why a conductor's surface field is σ/ϵ_0 rather than $\sigma/2\epsilon_0$.)*

16.31 ★★★ A uniformly charged solid sphere of radius R and charge $+Q$ is fixed in place, and a narrow tunnel is drilled through it along a diameter. A particle of mass m and charge $-q$ is released from rest at the surface, at the mouth of the tunnel.

a. Using equation (16.15), show that inside the sphere the force on the particle is $F = -(k_e qQ/R^3)\ r$, directed toward the center.
b. Show that the particle oscillates in simple harmonic motion, and find its period.
c. Compare with the "gravity tunnel" through a uniform Earth. Why do the two problems have the same form?

This page titled 16.E: Electric Fields (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17: GAUSS'S LAW

Chapter 16 found the electric fields of lines, rings, disks, planes, and spheres by adding up the contributions of countless tiny charges. The integrals worked, but only after symmetry had done most of the work, canceling components in pairs so that a single integral remained. This chapter turns that observation into a law. Instead of adding up fields charge by charge, we look at the field over a whole closed surface and ask how much of it passes outward through the surface. The answer, discovered by the German mathematician Carl Friedrich Gauss, is astonishingly simple: it depends only on the total charge inside.

That statement is **Gauss's law**. For charges at rest it says nothing that Coulomb's law does not, but it says it in a way that is often far more useful. When a charge distribution has enough symmetry, Gauss's law gives its field in two or three lines, with no integration at all. It also tells us, with almost no calculation, where the charge on a conductor must go and what the field just outside it must be. And it turns out to be the more fundamental of the two laws: Gauss's law remains true for moving charges and changing fields, where Coulomb's law fails, and it takes its place as the first of Maxwell's four equations of electromagnetism (Chapter 27).

Learning Objectives

- Use symmetry to determine the direction of an electric field and the variables it can depend on.
- Calculate the electric flux through an open or closed surface.
- State Gauss's law and explain why the flux through a closed surface depends only on the charge it encloses.
- Choose Gaussian surfaces that exploit cylindrical, planar, and spherical symmetry, and use them to find electric fields.
- Describe the properties of a conductor in electrostatic equilibrium, and find the charges and fields in and around conductors.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

17.1: The Importance of Symmetry
17.2: Electric Flux
17.3: Gauss's Law
17.4: Applications of Gauss's Law
17.5: Conductors in Electrostatic Equilibrium
17.S: Gauss's Law (Summary)
17.E: Gauss's Law (Problems)

This page titled 17: Gauss's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.1: The Importance of Symmetry

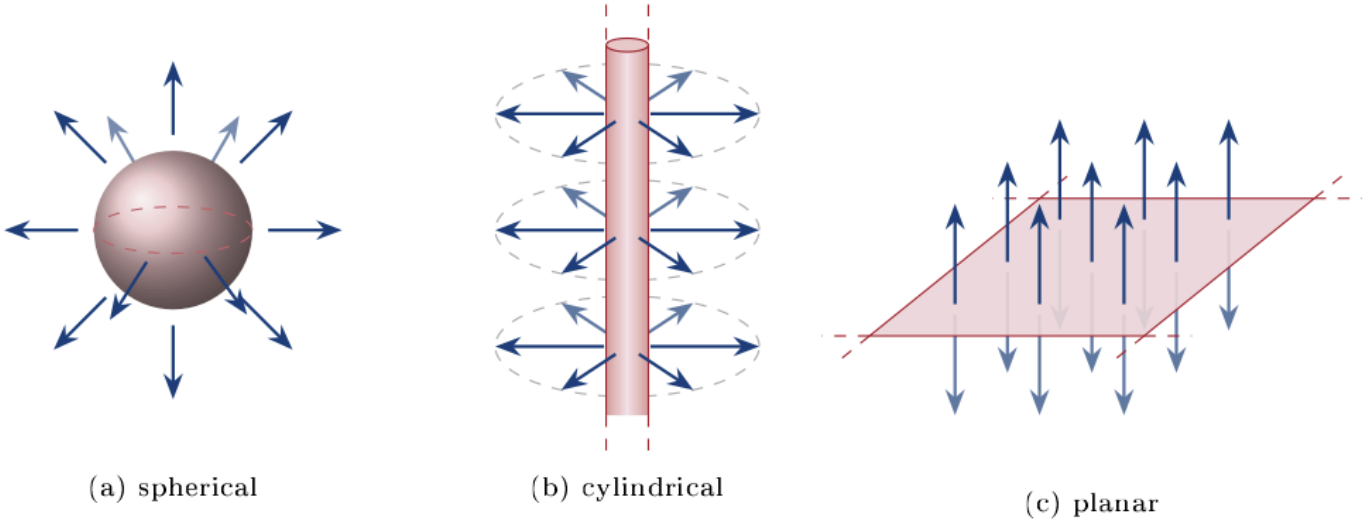


Figure 17.1. The three symmetries of this chapter, drawn for positive charge. (a) A spherically symmetric distribution has a radial field that depends only on the distance from the center. (b) An infinite line or cylinder has a field that points straight away from the axis, in every direction around it like the spokes of a wheel, and depends only on the distance from the axis. The arrows spread apart as they leave the axis. (c) An infinite plane has a field perpendicular to it that depends only on the distance from it. Every arrow on one side points the same way, so the arrows never spread apart.

•Figure description

Three panels, drawn in perspective, showing the field directions forced by symmetry. Panel a, spherical symmetry: a shaded ball with arrows pointing radially outward from its surface in every direction, with a dashed equator to show it is a sphere. Panel b, cylindrical symmetry: a long vertical cylinder of charge, its ends trailing off as dashed lines to show it continues forever. At three heights, a ring of six arrows fans out horizontally in all directions around the axis, like the spokes of a wheel, some pointing toward the viewer and some away. The arrow tips at each height lie on a dashed circle, so the arrows spread farther apart as they move away from the axis. Panel c, planar symmetry: a flat square sheet of charge seen from above at an angle, its edges trailing off as dashed lines to show it continues forever. A grid of nine arrows rises straight up from the top face and nine more point straight down from the bottom face. All the arrows on each side are parallel to one another and never spread apart.

Before calculating anything, it pays to ask what symmetry alone can tell us about a field. The key idea is simple: *if a charge distribution looks exactly the same after some operation, such as a rotation, a reflection, or a translation, then its field must look exactly the same after that operation too.* The field is produced by the charges and by nothing else, so it cannot have any feature that the charges lack. Three kinds of symmetry come up again and again: spherical, cylindrical, and planar (Figure 17.1).

Spherical symmetry. A point charge, a uniformly charged sphere, or a set of concentric spherical shells looks the same however it is rotated about its center. Its field must therefore point *radially*, straight toward or away from the center. Any sideways component would have to point in some particular direction, east rather than west, say, and nothing in the charge distribution picks out such a direction. For the same reason the field's magnitude can depend only on the distance r from the center, not on the direction. Such a distribution has **spherical symmetry**.

Cylindrical symmetry. An infinitely long line of charge, or an infinitely long uniformly charged cylinder, looks the same when rotated about its axis, when slid along its axis, and when flipped end for end. Its field must point radially away from (or toward) the axis, with a magnitude that depends only on the perpendicular distance r from the axis. A component along the axis is ruled out by the end-for-end flip, which would reverse it while leaving the charges unchanged. This is **cylindrical symmetry**.

Planar symmetry. An infinite plane of charge, or an infinite slab, looks the same when slid in any direction parallel to it and when rotated about any line perpendicular to it. Its field must be perpendicular to the plane, with a magnitude that depends only on the distance from it. If the charge is spread symmetrically about the plane's midpoint, reflecting through that midplane tells us that the field points away from the plane on both sides (or toward it on both sides), with equal magnitude at equal distances. This is **planar symmetry**.

Symmetry has told us the *direction* of the field everywhere, and which single variable its *magnitude* can depend on. What it cannot tell us is how the magnitude depends on that variable. That last step is where Gauss's law comes in. Chapter 16 used symmetry too, but only to cancel components inside an integral. Gauss's law uses it far more powerfully, to avoid the integral altogether.

Note: Symmetry must be exact

The arguments above need the symmetry to be complete. A rod $10\ \mathrm{cm}$ long is not cylindrically symmetric, because sliding it along its axis changes where its ends are. Its field near the middle is close to that of an infinite line (Section 16.3), but not exactly, and near the ends it is quite different. Gauss's law is *true* for the finite rod, as it is for every charge distribution, but without full symmetry it is not enough, on its own, to find the field. For such cases the integrals of Chapter 16, or a computer, are still needed.

Checkpoint 17.1

For each distribution, decide whether it has spherical, cylindrical, or planar symmetry, or none of these: (a) a uniformly charged solid ball; (b) a uniformly charged cube; (c) a very long, straight, charged copper pipe; (d) two concentric charged spherical shells carrying different charges; (e) a charged ring; (f) a very large, thick, uniformly charged sheet of plastic.

Answer

Answers to Checkpoints

Checkpoint 17.1. (a) Spherical. (b) None: a cube looks different along a face direction than along a diagonal, so the field is not the same at all points a given distance from the center. (c) Cylindrical, away from its ends. (d) Spherical: each shell is spherically symmetric about the common center, so their sum is too. (e) None. A ring has symmetry about its axis, but not enough for Gauss's law: no surface has a field of constant magnitude on it. (f) Planar, away from its edges.

This page titled 17.1: The Importance of Symmetry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.2: Electric Flux

The word *flux* comes from the Latin for “flow,” and a flowing fluid gives the clearest picture of what it measures. In Chapter 14 the volume of water per second passing through a pipe was Av, the area times the speed. If a small square loop is held in a stream, the volume per second passing through it is greatest when the loop faces the flow squarely, and zero when the loop is turned edge-on. At an angle in between, only the component of the velocity perpendicular to the loop carries water through it.

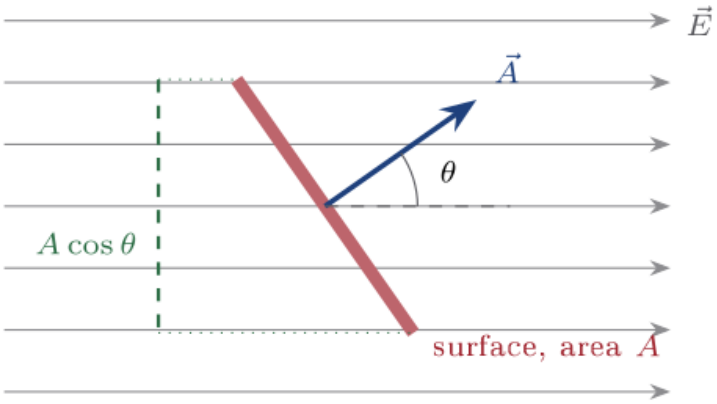


Figure 17.2. Flux through a flat surface, seen edge-on. The area vector $\mathbf{A}$ is perpendicular to the surface. The field lines that pass through the surface are exactly those that pass through its projection $A\cos\theta$ perpendicular to the field, so $\Phi_E = \mathbf{E}\cdot\mathbf{A} = EA\cos\theta$.

•Figure description

A flat surface of area A seen edge-on, tilted in a uniform electric field made of evenly spaced horizontal field lines pointing to the right. The area vector A starts at the middle of the surface and points perpendicular to it, at angle theta to the field direction. A dashed vertical line beside the surface shows its projection perpendicular to the field, of length A cos theta; only the field lines crossing this projection pass through the surface.

Electric flux applies the same idea to the electric field, even though nothing actually flows. To describe the orientation of a flat surface of area A, we use its **area vector** $\mathbf{A}$, a vector of magnitude A that points perpendicular to the surface. If the surface sits in a uniform field $\mathbf{E}$, and $\mathbf{A}$ makes an angle θ with $\mathbf{E}$ (Figure 17.2), the **electric flux** through the surface is

Key Equation: Electric flux through a flat surface in a uniform field

$$\Phi_E = \mathbf{E}\cdot\mathbf{A} = EA\cos\theta. \tag{17.1}$$

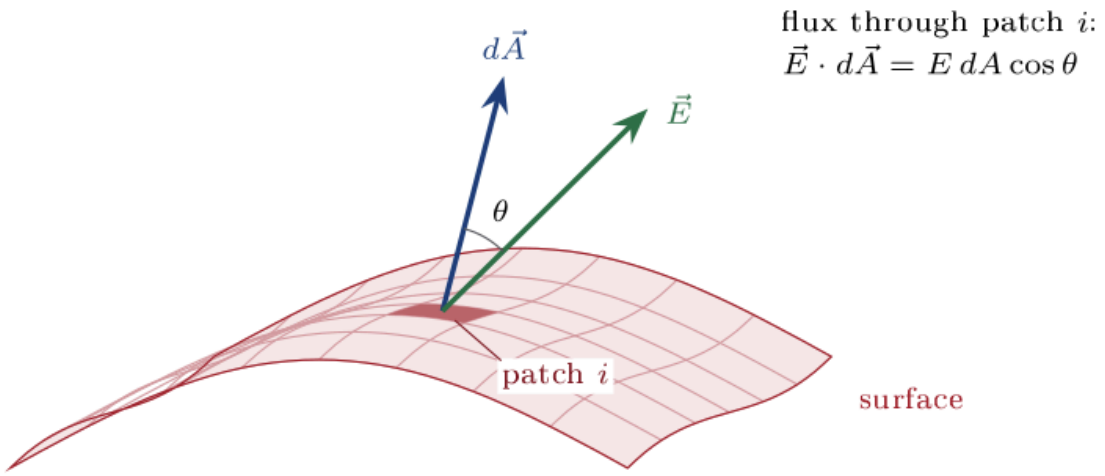


Figure 17.3. Flux through a curved surface. The surface is divided into small patches, each nearly flat. Patch i has an area vector $d\mathbf{A}$ perpendicular to it, and the field there makes an angle θ with $d\mathbf{A}$, so the patch contributes flux $\mathbf{E} \cdot d\mathbf{A} = E\ dA \cos\theta$. Adding the contributions of all the patches gives $\Phi_E = \int \mathbf{E} \cdot d\mathbf{A}$.

•Figure description

A curved surface drawn in perspective as a gently rolling sheet, divided by a grid of lines into many small patches. One patch, patch i, is shaded dark. From the center of that patch the area vector d A points perpendicular to the sheet, and the local electric field E points away from the patch at an angle theta to d A. A label states that the patch contributes flux E dot d A, which equals E d A cosine theta.

The SI unit of flux is the $\mathrm{N \cdot m^2/C}$. When the field is perpendicular to the surface ($\theta = 0$) the flux is EA, its largest value. When the field runs parallel to the surface ($\theta = 90^\circ$) the flux is zero, because no field passes through. You can read equation (17.1) two ways: as the perpendicular component of the field, $E\cos\theta$, times the whole area, or as the whole field times the area's projection $A\cos\theta$ onto a plane perpendicular to the field.

Flux has a vivid meaning in the field-line picture of Section 16.1. The number of field lines crossing a small area perpendicular to the field is drawn proportional to E times that area. So the electric flux through a surface is proportional to the *number of field lines that pass through it*. Tilting the surface lets fewer lines through, exactly as $\cos\theta$ says.

Curved surfaces and nonuniform fields. For a surface that is curved, or a field that varies from place to place, divide the surface into tiny patches, each small enough to be flat and to see a uniform field (Figure 17.3). Patch i has area vector $d\mathbf{A}$, of magnitude dA and perpendicular to the patch, and contributes flux $\mathbf{E} \cdot d\mathbf{A} = E\ dA \cos\theta$, where θ is the angle between $\mathbf{E}$ and $d\mathbf{A}$ at that patch. Adding them all up gives the general definition,

$$\Phi_E = \int \mathbf{E} \cdot d\mathbf{A}, \tag{17.2}$$

an integral over the surface.

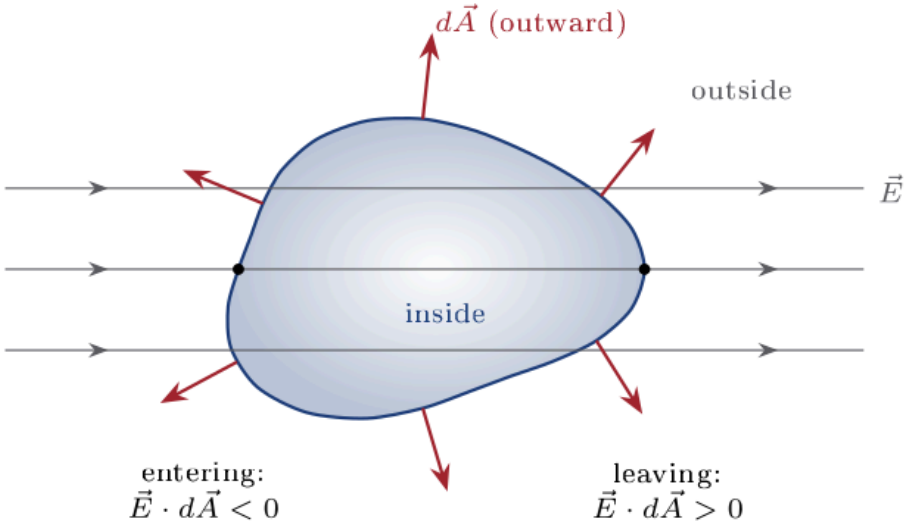


Figure 17.4. A closed surface encloses a volume completely, like the skin of a balloon or a potato. Every area vector $d\mathbf{A}$ is taken to point *outward*. Where a field line enters the surface, $\mathbf{E}$ points inward against $d\mathbf{A}$ and the flux is negative. Where it leaves, $\mathbf{E}$ points outward along $d\mathbf{A}$ and the flux is positive. In this uniform field every line that enters also leaves, so the net flux $\oint \mathbf{E} \cdot d\mathbf{A}$ is zero.

•Figure description

A closed surface shaped like a lumpy potato, shaded to look three-dimensional, sits in a uniform electric field of horizontal field lines pointing to the right. Short red area vectors d A stick straight out of the surface at several points all around it, every one pointing outward. On the left side, where a field line enters the surface, the field points inward while d A points outward, so E dot d A is negative. On the right side, where the same field line leaves, the field and d A both point outward, so E dot d A is positive.

Closed surfaces. A surface that encloses a volume completely, like the skin of a balloon, is a **closed surface** (Figure 17.4). It has a definite inside and outside, and anything leaving the inside must cross it. For a closed surface there is a natural choice of direction for every area vector: we always take $d\mathbf{A}$ to point *outward*. With that convention, field lines that leave the enclosed volume contribute positive flux, and field lines that enter it contribute negative flux. The net flux through a closed surface is written with a circle on the integral sign:

$$\Phi_E = \oint \mathbf{E} \cdot d\mathbf{A}. \tag{17.3}$$

Positive net flux means that more field lines leave the volume than enter it. Negative net flux means more enter than leave.

Example 17.1: Flux through a tilted rectangle

A rectangular sheet $20\ \text{cm}$ by $30\ \text{cm}$ sits in a uniform field of $4.0 \times 10^3\ \text{N/C}$. Its area vector makes an angle of 35° with the field. Find the flux through it. By how much does the flux change if the sheet is rotated so that it is parallel to the field?

Solution

The area is $A = (0.20)(0.30) = 0.060\ \text{m}^2$. From equation (17.1),

$$\Phi_E = EA\cos\theta = (4.0 \times 10^3\,\text{N/C})(0.060\,\text{m}^2)\cos 35^\circ = \boxed{2.0 \times 10^2\,\text{N} \cdot \text{m}^2/\text{C}}. \tag{17.2.1}$$

When the sheet is parallel to the field, its area vector is perpendicular to the field, $\theta = 90^\circ$, and the flux is zero. It decreases by the full $2.0 \times 10^2\ \text{N} \cdot \text{m}^2/\text{C}$. Note that the sign of an open surface's flux depends on which of its two normals we choose for $\mathbf{A}$. For an open surface that choice is ours to make. For a closed surface the outward convention makes it for us.

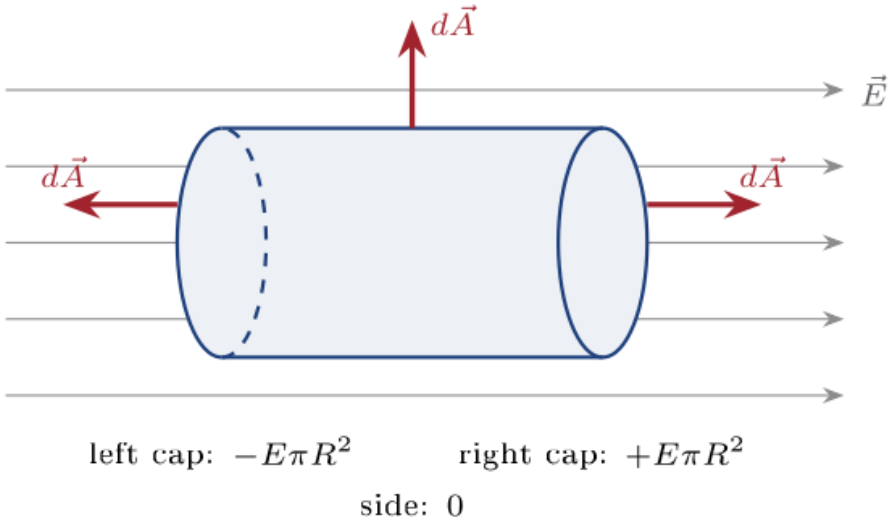


Figure 17.5. A closed cylinder in a uniform field. With outward area vectors, the left cap contributes negative flux, the right cap an equal positive flux, and the side none. The net flux is zero.

•Figure description

A closed cylinder lying on its side in a uniform field of horizontal field lines pointing to the right. The outward area vector of the left end cap points left, against the field; the outward area vector of the right end cap points right, along the field; an outward area vector on the curved side points straight up, perpendicular to the field. Field lines pass straight through the cylinder, entering at the left cap and leaving at the right cap.

Example 17.2: A closed surface in a uniform field

A closed cylinder of radius R and length L sits in a uniform field $\mathbf{E}$, with its axis parallel to the field (Figure 17.5). Find the net flux through its surface.

Solution

Split the closed surface into three parts: the left end cap, the right end cap, and the curved side. On the right cap, the outward normal points along $\mathbf{E}$, so $\theta = 0$ and the flux is $+E\pi R^2$. On the left cap, the outward normal points against $\mathbf{E}$, so $\theta = 180^\circ$ and the flux is $-E\pi R^2$. On the curved side, every outward normal is perpendicular to $\mathbf{E}$, so the flux is zero. The net flux is

$$\Phi_E = -E\pi R^2 + 0 + E\pi R^2 = \boxed{0}. \tag{17.2.2}$$

Every field line that enters through the left cap leaves through the right one. Nothing inside the cylinder creates or absorbs field lines, and, as the next section shows, that is precisely why the answer is zero.

Checkpoint 17.2

A cube sits in a uniform field that points along one of its edges. (a) Through which faces is the flux positive, negative, and zero? (b) What is the net flux through the cube? (c) Would your answer to (b) change if the cube were tilted at some odd angle to the field? Explain using field lines.

Answer

Answers to Checkpoints

Checkpoint 17.2. (a) Lines enter through the face the field points into (negative flux) and leave through the opposite face (positive flux). The four side faces are parallel to the field, so no lines cross them and their flux is zero. (b) Zero: the entering and leaving fluxes are equal in size, since the field is uniform and the two faces have the same area. (c) No. The cube encloses no charge, so every field line that enters it must also leave it, whatever its orientation. With the cube tilted, lines enter and leave through

different faces than before, but the net count is still zero.

This page titled 17.2: Electric Flux is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.3: Gauss's Law

Now place a point charge q at the center of a sphere of radius r, and find the flux through the sphere. On the sphere the field has the same magnitude everywhere, $E = k_e q/r^2$, and it points radially outward, parallel to the outward area vector of every patch. So $\mathbf{E} \cdot d\mathbf{A} = E\ dA$ everywhere, and

$$\Phi_E = \oint E\, dA = E \oint dA = \frac{k_e q}{r^2}\,(4\pi r^2) = 4\pi k_e q = \frac{q}{\epsilon_0}, \tag{17.3.1}$$

using $k_e = 1/4\pi\epsilon_0$. (Here is the payoff for writing the Coulomb constant with a 4π in it.) The radius has dropped out. The field weakens as $1/r^2$, and the area of the sphere grows as r^2, so every sphere centered on the charge, large or small, has the same flux through it. In the field-line picture this is obvious: every one of the lines leaving the charge crosses every one of the spheres.

That observation extends to a surface of *any* shape. However the closed surface bulges and folds, each field line leaving a charge inside it must eventually cross the surface to get out, and if it crosses back in somewhere, it must cross out again, so the net count for each line is exactly one crossing outward. The net flux through the surface is the same as through a sphere: q/ϵ_0. A charge *outside* the surface is different. Each of its field lines that enters the surface also leaves it, so its net contribution is zero. Finally, by the superposition principle, the field of many charges is the sum of their individual fields, and so the flux is the sum of their individual fluxes. Putting these facts together gives **Gauss's law**:

Key Equation: Gauss's law

$$\Phi_E = \oint \mathbf{E}\cdot d\mathbf{A} = \frac{q_{\text{enc}}}{\epsilon_0}. \tag{17.4}$$

The net electric flux through any closed surface equals the net charge enclosed by the surface, divided by ϵ_0. The closed surface can be real or imaginary. In applications it is usually imaginary, drawn wherever it is most convenient, and it is called a **Gaussian surface**.

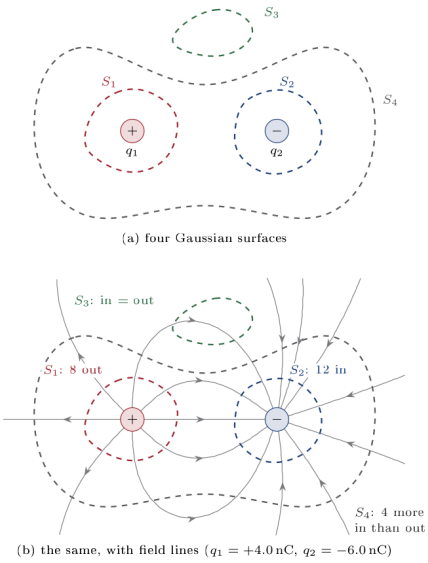


Figure 17.6. (a) Four Gaussian surfaces around two charges. The net flux through each depends only on the charge it encloses: q_1/ϵ_0 for S_1, q_2/ϵ_0 for S_2, zero for S_3, and $(q_1 + q_2)/\epsilon_0$ for S_4. (b) The same surfaces with the field lines drawn, eight lines for $q_1 = +4.0\ \mathrm{nC}$ and twelve for $q_2 = -6.0\ \mathrm{nC}$. Every line from q_1 crosses S_1 outward, and every line into q_2 crosses S_2 inward. Lines that pass through S_3 enter it and then leave it, so its net flux is zero. Lines that run off the edge of the drawing loop around far away and return to q_2. Some lines leave S_4 and come back in, but the four lines arriving from far away enter it and never leave, so its net flux is negative, matching its net enclosed charge of $-2.0\ \mathrm{nC}$.

•Figure description

Two panels showing the same arrangement. Panel a: a positive charge q sub 1 and a negative charge q sub 2 side by side, with four closed Gaussian surfaces drawn as dashed curves. Surface S sub 1 encloses only the positive charge; surface S sub 2 encloses only the negative charge; surface S sub 3, drawn above the charges, encloses neither; and the large surface S sub 4 encloses both charges. Panel b: the same charges and surfaces with the field lines drawn, for q sub 1 equal to plus 4 nanocoulombs and q sub 2 equal to minus 6 nanocoulombs. Eight field lines leave q sub 1, all crossing S sub 1 outward, and all eight curve around to end on q sub 2. Twelve lines end on q sub 2, all crossing S sub 2 inward: the eight from q sub 1 and four more that come in from far away on the right. Several of the arching lines from q sub 1 to q sub 2 pass through S sub 3, entering on one side and leaving on the other, so every line that enters S sub 3 also leaves it. The lines that arch high above the charges leave S sub 4 and come back in again, while the four lines from far away enter S sub 4 and never leave, so four more lines enter S sub 4 than leave it.

Three points about Gauss's law are easy to get wrong.

- The field $\mathbf{E}$ in the integral is the *total* field at each point of the surface, produced by all charges, inside and outside. Charges outside the surface do change the field on it. What they cannot change is the *net flux*, because their contributions in and out cancel.
- q_{enc} is the *net* charge inside, with signs. A surface enclosing $+q$ and $-q$ has zero net flux, even though the field on it is not zero.
- Zero net flux does not mean zero field. In Example 17.2 the field was the same everywhere and the net flux was still zero.

Example 17.3: Flux through four surfaces

In Figure 17.6, $q_1 = +4.0\ \mathrm{nC}$ and $q_2 = -6.0\ \mathrm{nC}$. Find the net flux through each of the four surfaces.

Solution

Apply equation (17.4) to each surface, counting only the charge inside:

$$\begin{aligned}\Phi_1 &= \frac{q_1}{\epsilon_0} = \frac{4.0\times10^{-9}}{8.85\times10^{-12}} = \boxed{+452\,\mathrm{N\cdot m^2/C}},\\ \Phi_2 &= \frac{q_2}{\epsilon_0} = \frac{-6.0\times10^{-9}}{8.85\times10^{-12}} = \boxed{-678\,\mathrm{N\cdot m^2/C}},\\ \Phi_3 &= \boxed{0},\\ \Phi_4 &= \frac{q_1+q_2}{\epsilon_0} = \frac{-2.0\times10^{-9}}{8.85\times10^{-12}} = \boxed{-226\,\mathrm{N\cdot m^2/C}}.\end{aligned} \tag{17.3.2}$$

The shapes and sizes of the surfaces never entered. Surface S_3 sits in a strong field and field lines certainly cross it, but every line that goes in comes back out.

Example 17.4: A charge at the center of a cube

A point charge $q = +6.0\ \mathrm{nC}$ sits at the center of a cube. Find the flux through one face of the cube.

Solution

Calculating the flux through a single face directly would mean integrating a field that varies in both magnitude and direction across the face. Symmetry avoids that. By Gauss's law the net flux through the whole cube is q/ϵ_0, and the charge sits at the center, so all six faces are equivalent and share it equally:

$$\Phi_{\text{face}} = \frac{q}{6\epsilon_0} = \frac{6.0\times10^{-9}}{6(8.85\times10^{-12})} = \boxed{113\,\mathrm{N\cdot m^2/C}}. \tag{17.3.3}$$

Example 17.5: Finding charge from a field

In a certain region the electric field is $\mathbf{E} = bx\ \hat{\imath}$, with $b = 3.0\times10^3\ \mathrm{N/(C\cdot m)}$. It points along x and grows in proportion to x. A cube of side $a = 0.20\ \mathrm{m}$ has one corner at the origin and its edges along the axes. How much charge is inside the cube?

Solution

The field is everywhere parallel to the x-axis, so only the two faces perpendicular to x have any flux. On the face at $x = 0$, $E = 0$ and there is no flux. On the face at $x = a$, $E = ba$, pointing along the outward normal, so the flux is $(ba)(a^2) = ba^3$. The net flux is

$$\Phi_E = ba^3 = (3.0\times10^3)(0.20)^3 = 24\,\mathrm{N\cdot m^2/C}, \tag{17.3.4}$$

and by Gauss's law,

$$q_{\text{enc}} = \epsilon_0\Phi_E = (8.85\times10^{-12})(24) = \boxed{2.1\times10^{-10}\,\mathrm{C}}. \tag{17.3.5}$$

A field that grows as it leaves the cube carries more flux out than in, and the excess must have been produced by charge inside. Here Gauss's law runs backward, from field to charge, which is how charge densities are often measured.

Note: Gauss's law for gravity

Nothing in the argument for Gauss's law used anything but the inverse-square law and superposition. The gravitational field $\mathbf{g}$ of Chapter 13 obeys both, so it has a Gauss's law of its own:

$$\oint \mathbf{g}\cdot d\mathbf{A} = -4\pi G M_{\text{enc}}, \tag{17.3.6}$$

with a minus sign because gravity points *toward* mass. The shell theorems that Newton proved by laborious integration follow from it in a few lines, exactly as the electric shell theorems do in Section 17.4.

Checkpoint 17.3

A closed surface encloses a net charge of zero. (a) Must the field be zero everywhere on the surface? (b) Must the net flux be zero? (c) A point charge $+q$ outside the surface is moved closer to it. Does the net flux change? Does the field on the surface change?

Answer

Answers to Checkpoints

Checkpoint 17.3. (a) No. Zero enclosed charge fixes only the net flux; the field on the surface can be large, produced by charges outside or by equal positive and negative charges inside. (b) Yes. By Gauss's law, net flux $= q_{\text{enc}}/\epsilon_0 = 0$. (c) The net flux does not change: it stays zero, because the charge is still outside. The field on the surface does change, growing stronger on the side nearer the charge, but every extra line entering the surface also leaves it.

This page titled 17.3: Gauss's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.4: Applications of Gauss's Law

Gauss's law holds for every closed surface around every charge distribution. To use it to *find* a field, though, we need to be able to pull E out of the integral, and that requires the symmetries of Section 17.1. The method is always the same.

Note: A strategy for Gauss's-law problems

1. Identify the symmetry, and use it to decide the direction of $\mathbf{E}$ and the variable its magnitude depends on.
2. Choose a Gaussian surface through the point where you want the field, shaped so that on each part of it $\mathbf{E}$ is either perpendicular to the surface with constant magnitude, or parallel to the surface so that it contributes no flux.
3. Write the flux, which will reduce to E times an area.
4. Find the charge enclosed by the surface. For a continuous distribution this may take a simple integral of the charge density.
5. Set the flux equal to $q_{\text{enc}}/\epsilon_0$ and solve for E.

Spheres

For a spherically symmetric distribution, the field is radial with magnitude $E(r)$, and the natural Gaussian surface is a sphere of radius r concentric with the charge. On it $\mathbf{E}$ is everywhere perpendicular to the surface and constant in magnitude, so the flux is $E(4\pi r^2)$, and Gauss's law gives

$$E(4\pi r^2) = \frac{q_{\text{enc}}}{\epsilon_0} \qquad \Longrightarrow \qquad E = \frac{q_{\text{enc}}}{4\pi\epsilon_0 r^2} = \frac{k_e\, q_{\text{enc}}}{r^2}. \tag{17.5}$$

Every spherically symmetric distribution obeys equation (17.5), and it contains a great deal.

A point charge. The Gaussian sphere encloses q, and $E = k_e q/r^2$. We have derived Coulomb's law from Gauss's law. The two contain the same physics for static charges, and either can be derived from the other.

The shell theorems. For a uniform spherical shell of charge Q and radius R, a Gaussian sphere with $r > R$ encloses Q, so $E = k_e Q/r^2$: outside, the shell acts as a point charge at its center. A Gaussian sphere with $r < R$ encloses nothing, so $E = 0$: inside the shell there is no field at all. These are the two shell theorems that Section 16.3 borrowed from Newton, now proved in two lines.

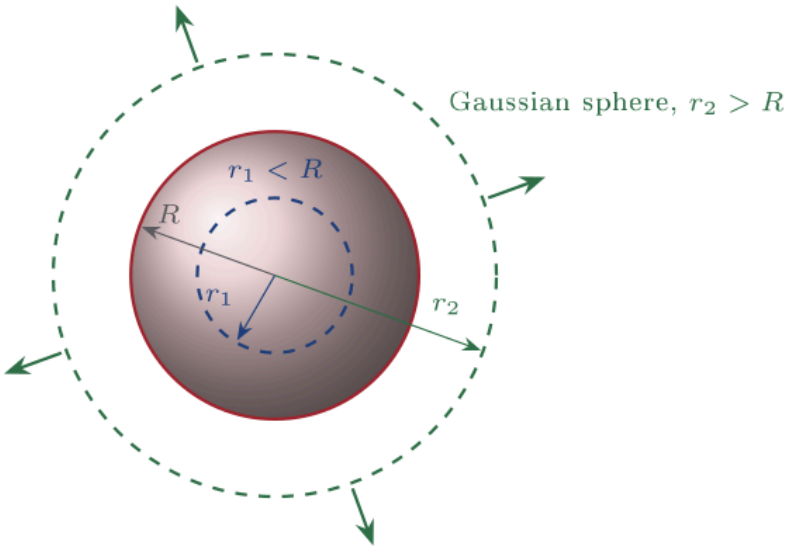


Figure 17.7. Spherical Gaussian surfaces for a uniformly charged solid sphere. A Gaussian sphere outside encloses all the charge. One inside, of radius r_1, encloses only the fraction $(r_1/R)^3$ of it, which is why the interior field grows linearly with r.

•Figure description

A uniformly charged solid sphere of radius R, shaded, with two dashed concentric Gaussian spheres. The inner Gaussian sphere, of radius r sub 1 less than R, lies inside the charged sphere and encloses only part of its charge. The outer Gaussian sphere, of radius r sub 2 greater than R, encloses all of it. Radial field arrows are drawn on the outer Gaussian sphere.

A uniformly charged solid sphere. For charge Q spread uniformly through a sphere of radius R (Figure 17.7), a Gaussian sphere outside encloses Q, and a Gaussian sphere inside, at radius r, encloses the fraction $(r/R)^3$ of it. Equation (17.5) then gives

$$E = \frac{k_e Q}{r^2} \quad (r \geq R), \qquad E = \frac{k_e Q r^3/R^3}{r^2} = \frac{k_e Q}{R^3}\, r \quad (r \leq R), \tag{17.4.1}$$

exactly equation (16.15).

Gauss's law does more than reproduce old results. It handles distributions whose integrals would be miserable, as long as they are spherically symmetric.

Example 17.6: A sphere whose charge density grows outward

An insulating sphere of radius R carries charge with density $\rho = \rho_0 r/R$, increasing from zero at the center to ρ_0 at the surface. Find the total charge and the field inside and outside the sphere.

Solution

The charge is not uniform, but it depends only on r, so the distribution is spherically symmetric and equation (17.5) applies. We only need q_{enc}. Divide the sphere into thin shells of radius r' and thickness dr', each with volume $4\pi r'^2\ dr'$. The charge inside radius r is

$$q_{\text{enc}}(r) = \int_0^r \rho_0 \frac{r'}{R}\, 4\pi r'^2\, dr' = \frac{4\pi\rho_0}{R} \cdot \frac{r^4}{4} = \frac{\pi\rho_0 r^4}{R}, \tag{17.4.2}$$

and the total charge is $Q = q_{\text{enc}}(R) = \boxed{\pi\rho_0 R^3}$.

Inside, $r \leq R$,

$$E = \frac{q_{\text{enc}}}{4\pi\epsilon_0 r^2} = \frac{\pi\rho_0 r^4/R}{4\pi\epsilon_0 r^2} = \boxed{\frac{\rho_0\, r^2}{4\epsilon_0 R}}. \tag{17.4.3}$$

Outside, $r \geq R$, $E = \boxed{k_e Q/r^2}$, as for any spherical distribution. Check the surface: the inside formula gives $\rho_0 R/4\epsilon_0$, and the outside formula gives $\pi\rho_0 R^3/(4\pi\epsilon_0 R^2) = \rho_0 R/4\epsilon_0$. They agree. Inside, the field now grows as r^2 rather than r, because more of the charge is concentrated toward the outside.

Checkpoint 17.4

A point charge $+q$ sits at the center of a thin spherical shell that carries a charge $-2q$ spread uniformly. (a) What is the field inside the shell, at distance r from the center? (b) What is the field outside the shell? (c) Sketch a few field lines. In which region do they point outward, and in which inward?

Answer

Cylinders

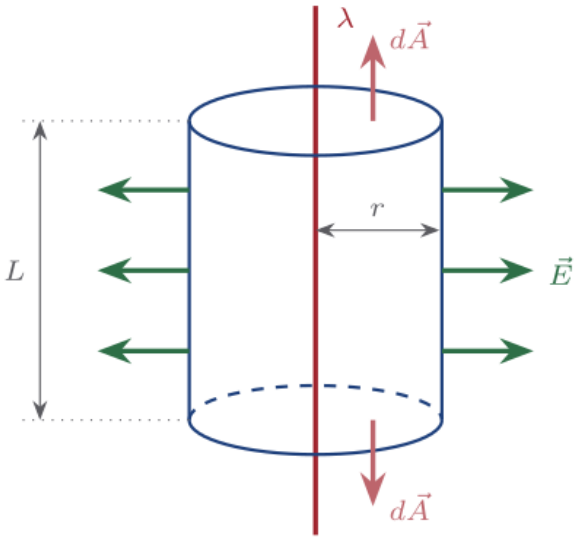


Figure 17.8. A cylindrical Gaussian surface around an infinite line of charge. The field is perpendicular to the curved side and the same everywhere on it. The end caps contribute no flux because the field runs parallel to them.

•Figure description

A long vertical line of charge with a coaxial Gaussian cylinder of radius r and length L drawn around it. On the curved side of the cylinder, arrows show the field pointing radially outward, perpendicular to the surface. On the top and bottom end caps, arrows show the outward area vectors pointing up and down, perpendicular to the field, so the caps carry no flux.

An infinite line of charge. A long straight line carries uniform linear charge density λ. By cylindrical symmetry its field points radially away from the line and has a magnitude $E(r)$ that depends only on the distance r from it. For the Gaussian surface, take a closed cylinder of radius r and length L, coaxial with the line (Figure 17.8).

On the curved side of the cylinder, $\mathbf{E}$ points radially outward, along the outward normal, and has the same magnitude $E(r)$ everywhere, so the flux through the side is E times its area, $E(2\pi rL)$. On the two flat end caps the field is parallel to the surface, so their flux is zero. The cylinder encloses a length L of the line, and so a charge λL. Gauss's law gives

$$E(2\pi rL) = \frac{\lambda L}{\epsilon_0} \qquad \Longrightarrow \qquad E = \frac{\lambda}{2\pi\epsilon_0 r}. \tag{17.4.4}$$

This is equation (16.10), which took an integral and a limit in Chapter 16. The arbitrary length L canceled, as it had to.

A charged cylindrical shell. Suppose instead that the charge λ per unit length is spread uniformly over the surface of a long, thin-walled cylinder of radius R, a charged pipe. Outside it ($r > R$) the same Gaussian cylinder encloses the same charge λL and gives the same field, $\lambda/2\pi\epsilon_0 r$: from outside, a charged pipe is indistinguishable from a charged line along its axis. Inside ($r < R$) the Gaussian cylinder encloses no charge at all, so $E(2\pi rL) = 0$ and the field is *zero* everywhere inside the pipe.

Example 17.7: A charged insulating rod

A long, solid plastic rod of radius $R = 1.0\ \text{cm}$ carries a uniform volume charge density $\rho = 2.0 \times 10^{-5}\ \text{C/m}^3$. Find the electric field (a) $0.50\ \text{cm}$ from the axis, inside the rod; (b) at its surface; and (c) $3.0\ \text{cm}$ from the axis.

Solution

The rod is cylindrically symmetric, so we use a coaxial Gaussian cylinder of radius r and length L. In every case the flux is $E(2\pi rL)$. Only the enclosed charge differs.

Inside, $r < R$, the Gaussian cylinder encloses the volume $\pi r^2 L$, and so charge $\rho\pi r^2 L$:

$$E(2\pi rL) = \frac{\rho\pi r^2 L}{\epsilon_0} \qquad \Longrightarrow \qquad E = \frac{\rho r}{2\epsilon_0}. \tag{17.4.5}$$

Outside, $r > R$, it encloses the whole cross-section, charge $\rho\pi R^2 L$, which is the same as a line with $\lambda = \rho\pi R^2$:

$$E = \frac{\rho R^2}{2\epsilon_0 r}. \tag{17.4.6}$$

(a) At $r = 0.0050\ \text{m}$,

$$E = \frac{(2.0 \times 10^{-5})(0.0050)}{2(8.85 \times 10^{-12})} = \boxed{5.6 \times 10^3\ \text{N/C}}. \tag{17.4.7}$$

(b) At $r = R = 0.010\ \text{m}$, either formula gives $\boxed{1.13 \times 10^4\ \text{N/C}}$.

(c) At $r = 0.030\ \text{m}$,

$$E = \frac{(2.0 \times 10^{-5})(0.010)^2}{2(8.85 \times 10^{-12})(0.030)} = \boxed{3.8 \times 10^3\ \text{N/C}}. \tag{17.4.8}$$

The field grows linearly inside the rod, peaks at its surface, and falls off as $1/r$ outside, the cylindrical counterpart of the uniformly charged solid sphere of Section 17.4 (compare Figure 16.12). All three fields point radially away from the axis.

Planes

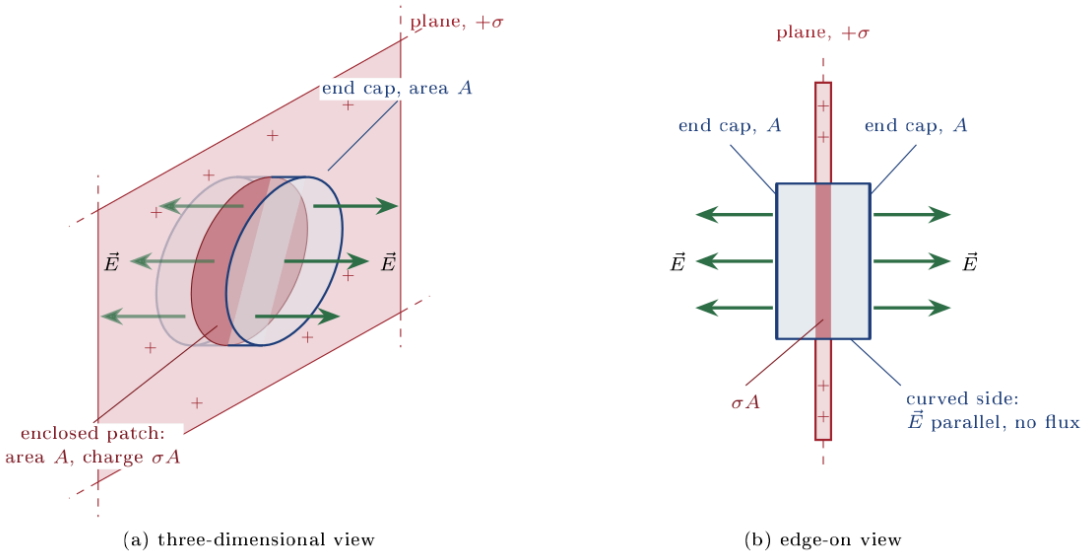


Figure 17.9. A pillbox Gaussian surface through an infinite plane of charge. (a) In three dimensions the pillbox passes straight through the plane, and its two end caps, each of area A, are parallel to the plane. The plane cuts out a patch of area A inside the pillbox, carrying charge σA. (b) Seen edge-on, the end caps and the plane appear as parallel lines. Flux leaves through both end caps and none through the curved side, so $2EA = \sigma A/\epsilon_0$.

•Figure description

Two panels showing a pillbox Gaussian surface through an infinite plane of positive charge. Panel a, a three-dimensional view: the charged plane is drawn as a large, slightly transparent sheet seen at an angle, its edges trailing off as dashed lines to show it continues forever. A short, wide cylinder, a pillbox shaped like a coin, passes straight through the sheet with its axis perpendicular to it. Its two circular end caps, each of area A, are parallel to the sheet, one on each side at equal distances. Where the sheet cuts through the pillbox, a circular patch of the sheet, drawn in the same perspective as the sheet and the caps, is shaded darker: this is the enclosed charge, of area A and charge sigma A. Field arrows point straight out of both end caps, away from the plane. Panel b, the same arrangement seen edge-on: the plane is a vertical band, and the pillbox is a rectangle straddling it, with its two end caps drawn as vertical lines parallel to the band. The short piece of the band inside the rectangle is shaded as the enclosed charge. Field arrows leave through both end caps, perpendicular to them, and run parallel to the top and bottom sides, which represent the curved side of the pillbox.

An infinite plane of charge. An infinite plane carries uniform surface charge density σ. By planar symmetry, its field is perpendicular to the plane and points away from it on both sides, with a magnitude that can depend only on the distance from the plane. For the Gaussian surface take a short closed cylinder, often called a **pillbox**, that pierces the plane at right angles, with end caps of area A an equal distance from the plane on either side (Figure 17.9).

On each end cap the field points outward, along the outward normal, with magnitude E, so each cap contributes flux EA. The field runs parallel to the curved side, which contributes nothing. The pillbox encloses a patch of the plane of area A, and so charge σA . Therefore

$$2EA = \frac{\sigma A}{\epsilon_0} \qquad \Longrightarrow \qquad E = \frac{\sigma}{2\epsilon_0}, \tag{17.4.9}$$

which is equation (16.13). Notice that we never had to specify how far the end caps are from the plane. Gauss's law does not merely tell us the field at one distance: it shows that the field is the same at every distance, a result that was a surprise in Chapter 16. The field of two oppositely charged plates, σ/ϵ_0 between them and zero outside (equation (16.14)), follows from this result by superposition, just as before.

Example 17.8: A thick charged slab

A large, flat plastic slab of thickness $d = 2.0\ \text{cm}$ carries a uniform volume charge density $\rho = 3.0 \times 10^{-6}\ \text{C/m}^3$. Find the electric field (a) $0.50\ \text{cm}$ from the slab's midplane, inside it, and (b) outside the slab.

Solution

The slab has planar symmetry about its midplane, so the field points away from the midplane on both sides and depends only on the distance x from it. Use a pillbox centered on the midplane with end caps of area A at $\pm x$. The flux is $2EA$, as for the thin plane.

Inside, $x < d/2$, the pillbox encloses a thickness $2x$ of the slab, charge $\rho(2xA)$:

$$2EA = \frac{2\rho x A}{\epsilon_0} \qquad \Longrightarrow \qquad E = \frac{\rho x}{\epsilon_0}. \tag{17.4.10}$$

At $x = 0.0050\ \text{m}$,

$$E = \frac{(3.0 \times 10^{-6})(0.0050)}{8.85 \times 10^{-12}} = \boxed{1.7 \times 10^3\,\text{N/C}}. \tag{17.4.11}$$

Outside, $x > d/2$, the pillbox encloses the full thickness, charge ρdA:

$$E = \frac{\rho d}{2\epsilon_0} = \frac{(3.0 \times 10^{-6})(0.020)}{2(8.85 \times 10^{-12})} = \boxed{3.4 \times 10^3\,\text{N/C}}, \tag{17.4.12}$$

the same at every distance outside. From outside, the slab acts exactly like a thin plane with $\sigma = \rho d$. The field is zero at the midplane, grows linearly to the surfaces, and is constant beyond them.

Answers to Checkpoints

Checkpoint 17.4. (a) A Gaussian sphere of radius r inside the shell encloses only $+q$, so $E = k_e q/r^2$, pointing outward, exactly as if the shell were absent. (b) A Gaussian sphere outside encloses $q - 2q = -q$, so $E = k_e q/r^2$, pointing *inward*, as if a charge $-q$ sat at the center. (c) Inside, lines point outward from $+q$ and end on the shell. Outside, lines come in from far away and end on the shell. The shell's $-2q$ absorbs all the lines from $+q$ plus an equal number more from outside.

This page titled 17.4: Applications of Gauss's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.5: Conductors in Electrostatic Equilibrium

A conductor, as we saw in Section 15.1, contains free electrons that move in response to any electric force. When charge is placed on a conductor, or a conductor is brought into a field, those electrons move, but only for a very short time, typically far less than a microsecond. They rearrange themselves until the net force on every free electron is zero, and then they stop. A conductor in this settled state, with no net motion of charge, is in **electrostatic equilibrium**. Gauss's law tells us a great deal about it, with almost no calculation.

1. The field inside the conducting material is zero. If there were a field anywhere inside the metal, the free electrons there would feel a force and move, and the conductor would not be in equilibrium. So in equilibrium $\mathbf{E} = 0$ everywhere within the material. This is not a special property of the charges' arrangement. It is *what equilibrium means* for a conductor, and the free charges arrange themselves, whatever external field is applied, until it is true.

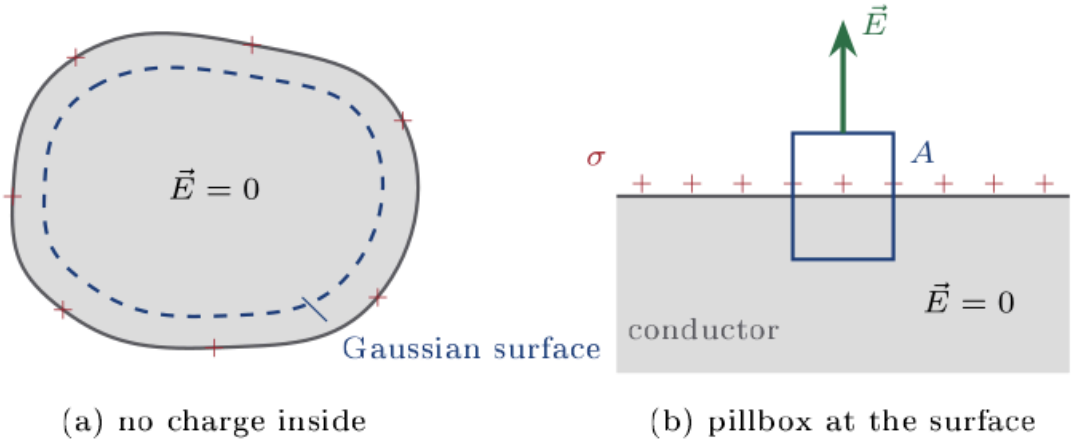


Figure 17.10. (a) A Gaussian surface just inside a conductor lies where $\mathbf{E} = 0$, so it encloses no net charge: excess charge lives on the surface. (b) A pillbox straddling the surface. Flux leaves only through the outer cap, giving $E = \sigma/\epsilon_0$ just outside.

•Figure description

Two panels showing the edge of a charged conductor. In panel a, the conductor is a shaded irregular blob with plus signs along its boundary, and a dashed Gaussian surface is drawn just inside the boundary, entirely within the metal, where the field is zero. In panel b, a small section of the conductor's surface is shown with plus charges along it; a small pillbox straddles the surface, with one end cap outside in the air and one inside the metal. The field E points straight out of the surface through the outer end cap, and inside the metal E equals zero.

2. Any excess charge lies on the surface. Draw a Gaussian surface just inside the surface of the conductor, entirely within the metal (Figure 17.10a). The field is zero at every point of it, so the flux through it is zero, and by Gauss's law it encloses no net charge. That surface can be drawn as close to the real surface as we like, so there can be no net charge anywhere in the interior. Any excess charge on a conductor must sit on its surface. This is the result Section 16.3 promised.

3. Just outside, the field is perpendicular to the surface. If the field just outside had a component parallel to the surface, it would push the surface charges along the surface, and they would move. In equilibrium, the field at the surface of a conductor has no parallel component. Field lines leave a charged conductor at right angles to its surface.

4. Just outside, the field is σ/ϵ_0. Let the surface charge density at some point of the surface be σ. Draw a small pillbox with one end cap of area A just outside the surface and the other just inside, in the metal (Figure 17.10b). The field is zero on the inner cap, and it is perpendicular to the surface, so it runs parallel to the pillbox's side. Only the outer cap carries flux, EA. The pillbox encloses charge σA, so $EA = \sigma A/\epsilon_0$:

Key Equation: Field just outside a conductor

$$E = \frac{\sigma}{\epsilon_0}. \tag{17.6}$$

This is *twice* the field of an isolated plane of charge with the same σ, equation (16.13). The difference is that the thin plane sends flux out through both caps of its pillbox, while at a conductor's surface the field inside is zero and all the flux goes out one side. Put another way: the patch of charge right next to the point contributes $\sigma/2\epsilon_0$ on each side, pointing away from it. All the *other* charges on the conductor contribute a further $\sigma/2\epsilon_0$ outward just outside, which adds, and $\sigma/2\epsilon_0$ outward just inside, which exactly cancels the patch's inward field there. The result is σ/ϵ_0 outside and zero inside.

5. Charge crowds toward sharp points. On a conducting sphere the charge spreads uniformly, by symmetry. On an irregular conductor it does not: the surface charge density, and with it the field just outside, is largest where the surface curves most sharply. We will be able to show why in Chapter 18. The consequence is that sharp points on a charged conductor can produce fields strong enough to ionize the surrounding air, which then glows and leaks charge away. This **corona discharge** is the faint glow sometimes seen on ships' masts and aircraft wingtips in stormy weather, and it is the principle behind the lightning rod.

Cavities and shielding. Suppose a conductor has an empty hole in it, a cavity containing no charge. A Gaussian surface drawn within the metal around the cavity has zero flux, so the cavity's walls carry no net charge. A more careful argument, which needs the electric potential of Chapter 18, shows that they carry no charge at all and that the field inside an empty cavity is exactly zero, *whatever* fields exist outside the conductor. A closed conducting shell therefore shields its interior from external electric fields. Such a shell is called a **Faraday cage**, after Faraday, who sat inside a large one while it was charged until sparks flew from its outside. Even a wire mesh works well, provided its holes are small. This is why the occupants of a car struck by lightning are usually unharmed (it is the metal body, not the rubber tires, that protects them), why a cell phone loses its signal in an elevator, and why the window of a microwave oven is covered with a fine metal mesh.

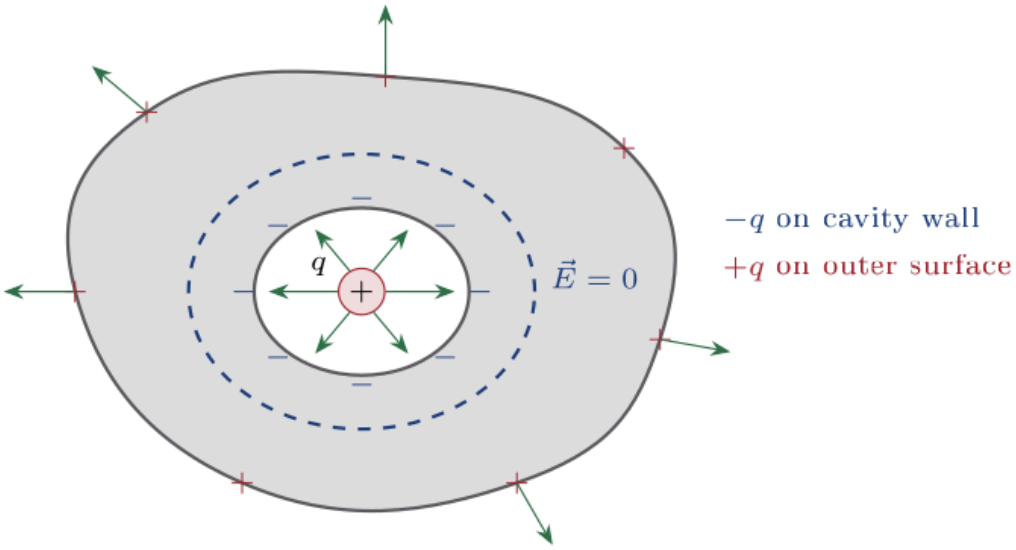


Figure 17.11. A neutral conductor with a charge q in its cavity. Because $\mathbf{E} = 0$ in the metal, the dashed Gaussian surface encloses no net charge, so $-q$ is induced on the cavity wall, and $+q$ on the outer surface keeps the conductor neutral.

•Figure description

A neutral conductor, drawn as a shaded blob, with an empty cavity inside it that contains a positive point charge q. Negative signs line the wall of the cavity, representing an induced charge of minus q, and positive signs line the outer surface of the conductor, representing an induced charge of plus q. A dashed Gaussian surface runs through the metal between the cavity and the outer surface, where the field is zero. Field lines run from the charge q outward to the negative charges on the cavity wall, and outward from the positive charges on the outer surface.

If the cavity *does* contain a charge q, the conductor responds (Figure 17.11). A Gaussian surface within the metal must enclose zero net charge, so a charge $-q$ gathers on the cavity wall. The conductor's total charge is fixed, so if it was neutral, a charge $+q$ must appear on its outer surface. The outside world therefore still "sees" the charge in the cavity, through the charge it induces on the outer surface. The conductor shields the inside from the outside, but not the outside from the inside, unless the outer

surface is grounded, which lets the $+q$ drain away.

Example 17.9: A charge inside a conducting shell

A point charge $q = +2.0\ \text{nC}$ sits at the center of a thick conducting spherical shell with inner radius $R_1 = 10\ \text{cm}$ and outer radius $R_2 = 15\ \text{cm}$. The shell carries a net charge of $-5.0\ \text{nC}$. (a) How is the shell's charge divided between its inner and outer surfaces? (b) Find the field at $r = 5.0\ \text{cm}$, $r = 12\ \text{cm}$, and $r = 20\ \text{cm}$.

Solution

(a) A Gaussian sphere of radius between R_1 and R_2 lies in the metal, where $E = 0$, so it encloses zero net charge. The inner surface must therefore carry $\boxed{-2.0\,\text{nC}}$, canceling the point charge. The remaining $-5.0\ \text{nC} - (-2.0\ \text{nC}) = \boxed{-3.0\,\text{nC}}$ sits on the outer surface. By symmetry, each is spread uniformly.

(b) Everything is spherically symmetric, so equation (17.5) applies in every region.

At $r = 0.050\ \text{m}$, inside the cavity, only the point charge is enclosed:

$$E = \frac{k_e q}{r^2} = \frac{(8.99 \times 10^9)(2.0 \times 10^{-9})}{(0.050)^2} = \boxed{7.2 \times 10^3\,\text{N/C}},\ \text{outward}. \tag{17.5.1}$$

At $r = 0.12\ \text{m}$, in the metal, $\boxed{E = 0}$.

At $r = 0.20\ \text{m}$, outside, the enclosed charge is $+2.0 - 5.0 = -3.0\ \text{nC}$:

$$E = \frac{k_e |q_{\text{enc}}|}{r^2} = \frac{(8.99 \times 10^9)(3.0 \times 10^{-9})}{(0.20)^2} = \boxed{6.7 \times 10^2\,\text{N/C}},\ \text{inward}. \tag{17.5.2}$$

Notice that the field in the cavity would be unchanged if the shell carried any other net charge, or if charges were brought near the outside of it. The shell's outer surface, and the world beyond, simply do not reach in.

Checkpoint 17.5

(a) A hollow metal sphere is given a large positive charge. Where does the charge go, and what is the field at the center of the hollow? (b) A student says: "The field just outside a conductor is σ/ϵ_0, but the field of a charged plane is $\sigma/2\epsilon_0$. A small patch of a large conductor looks just like a plane, so one of these must be wrong." Resolve the student's difficulty.

Answer

Answers to Checkpoints

Checkpoint 17.5. (a) All of the charge goes to the outer surface, as far apart as it can get; none resides on the inner surface of the empty hollow. The field at the center, and everywhere in the hollow, is zero. (b) Both formulas are right; they describe different situations. A lone charged sheet has field lines on both sides, so its field, $\sigma/2\epsilon_0$, points away from it on both sides. At a conductor's surface, the small patch contributes $\sigma/2\epsilon_0$ on each side, and the rest of the conductor's charge contributes a field that exactly cancels the patch's field inside the metal (where E must be zero) and so *adds* to it outside. The field just outside is $\sigma/2\epsilon_0 + \sigma/2\epsilon_0 = \sigma/\epsilon_0$: all the field lines leave through one side.

This page titled 17.5: Conductors in Electrostatic Equilibrium is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.S: Gauss's Law (Summary)

Summary

- **Symmetry** fixes the direction of a field and the single variable its magnitude can depend on. Spherically symmetric fields are radial and depend on r; cylindrically symmetric fields point away from the axis and depend on the distance from it; planar fields are perpendicular to the plane and depend on the distance from it.
- The **electric flux** through a surface is $\Phi_E = \int \mathbf{E} \cdot d\mathbf{A}$, which is $EA\cos\theta$ for a flat surface in a uniform field. It measures the number of field lines passing through the surface. For a closed surface, $d\mathbf{A}$ points outward, so outgoing lines count positive and incoming lines negative.
- **Gauss's law**: the net flux through any closed surface equals the enclosed charge divided by ϵ_0, $\oint \mathbf{E} \cdot d\mathbf{A} = q_{\text{enc}}/\epsilon_0$. Charges outside contribute to the field on the surface but not to the net flux.
- With a Gaussian surface matched to the symmetry, Gauss's law gives fields directly: $k_e q_{\text{enc}}/r^2$ for any spherical distribution, $\lambda/2\pi\epsilon_0 r$ for an infinite line, and $\sigma/2\epsilon_0$ for an infinite plane. It proves the shell theorems, recovers Coulomb's law, and handles nonuniform distributions such as $\rho \propto r$.
- A conductor in **electrostatic equilibrium** has $\mathbf{E} = 0$ inside its material; its excess charge lies on its surface; the field just outside is perpendicular to the surface with magnitude σ/ϵ_0; and charge concentrates at sharp points. An empty cavity in a conductor is field-free (a Faraday cage). A charge q in a cavity induces $-q$ on the cavity wall.

Key Equations

Electric flux

$$\Phi_E = \mathbf{E} \cdot \mathbf{A} = EA\cos\theta, \qquad \Phi_E = \int \mathbf{E} \cdot d\mathbf{A} \tag{17.S.1}$$

Gauss's law

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{q_{\text{enc}}}{\epsilon_0} \tag{17.S.2}$$

Spherically symmetric distribution

$$E = \frac{q_{\text{enc}}}{4\pi\epsilon_0 r^2} \tag{17.S.3}$$

Infinite line, and infinite plane

$$E = \frac{\lambda}{2\pi\epsilon_0 r}, \qquad E = \frac{\sigma}{2\epsilon_0} \tag{17.S.4}$$

Field just outside a conductor

$$E = \frac{\sigma}{\epsilon_0} \tag{17.S.5}$$

This page titled 17.S: Gauss's Law (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

17.E: Gauss's Law (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $k_e = 8.99 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$, $\epsilon_0 = 8.85 \times 10^{-12}\ \text{C}^2/(\text{N}\cdot\text{m}^2)$, and $e = 1.602 \times 10^{-19}\ \text{C}$.

Section 17.1 — The Importance of Symmetry

17.1 ★ *[Conceptual.]* Use a symmetry argument to explain why the field of an infinite line of charge cannot have a component parallel to the line.

17.2 ★★ *[Conceptual.]* A very long, straight wire carries a uniform charge. A student argues that its field might circle around the wire, like water swirling around a drain, while still depending only on the distance from the wire. What symmetry operation rules this out? *(Hint: reflect the wire in a mirror that contains the wire.)*

17.3 ★★ *[Conceptual.]* Gauss's law is true for a uniformly charged cube. Explain why it is nevertheless of little use in finding the field of the cube, and why it *is* useful for finding the field far from the cube.

Section 17.2 — Electric Flux

17.4 ★ A flat circular loop of radius $12\ \text{cm}$ sits in a uniform field of $6.0 \times 10^3\ \text{N/C}$. Find the flux through it when its area vector is (a) parallel to the field, (b) at 60° to the field, and (c) perpendicular to the field.

17.5 ★ *[Real-world: atmospheric physics.]* The fair-weather field near the ground is about $150\ \text{N/C}$, pointing straight down. Find the flux through a horizontal $2.0\ \text{m}$ by $3.0\ \text{m}$ flat roof, taking the area vector to point upward.

17.6 ★★ A uniform field $\mathbf{E} = (3.0 \times 10^3\ \text{N/C})\ \hat{\imath} + (4.0 \times 10^3\ \text{N/C})\ \hat{\jmath}$ passes through a square of side $0.10\ \text{m}$ lying in the yz-plane. Find the magnitude of the flux through the square. Repeat for a square of the same size in the xz-plane and in the xy-plane.

17.7 ★★ A hemispherical bowl of radius R sits with its open face perpendicular to a uniform field $\mathbf{E}$. Show that the flux through the curved surface of the bowl is $E\pi R^2$ in magnitude. *(Hint: the bowl and the flat disk across its opening together form a closed surface enclosing no charge.)*

17.8 ★★ A closed triangular box (a wedge) sits in a uniform horizontal field E. Its vertical back face is perpendicular to the field, which enters the box through that face, its bottom is horizontal, and its sloping top face is inclined at 30° to the horizontal. The sides are triangles parallel to the field. If the vertical face has area A, find the flux through each of the five faces and show that the total is zero.

Section 17.3 — Gauss's Law

17.9 ★ A closed surface encloses charges of $+3.0\ \text{nC}$, $-7.0\ \text{nC}$, and $+1.5\ \text{nC}$. A charge of $+5.0\ \text{nC}$ sits just outside it. Find the net flux through the surface.

17.10 ★ The net flux through a closed surface is $-850\ \text{N}\cdot\text{m}^2/\text{C}$. How much charge does it enclose, and how many excess electrons does that represent?

17.11 ★★ A point charge of $+8.0\ \text{nC}$ sits at one corner of a cube. Find the flux through each face of the cube. *(Hint: imagine seven more identical cubes arranged so that the charge sits at the center of a larger cube.)*

17.12 ★★ *[Conceptual.]* A point charge sits inside a closed spherical surface, but not at its center. (a) Is the flux through the sphere different from what it would be with the charge at the center? (b) Is the field on the sphere the same at every point? (c) Could you use Gauss's law to find the field on the sphere? Explain.

17.13 ★★★ In a region of space the field points in the $+x$ direction and has magnitude $E = cx^2$, where $c = 2.0 \times 10^4$

$\mathrm{N/(C \cdot m^2)}$. Find the net charge inside a cube of side $0.30\ \mathrm{m}$ whose faces lie at $x = 0.10\ \mathrm{m}$ and $x = 0.40\ \mathrm{m}$ (and whose other faces are parallel to the x-axis).

Section 17.4 — Applications of Gauss's Law

17.14 ★ A long straight wire carries $-4.0\ \mathrm{nC/m}$. Use Gauss's law to find the field $5.0\ \mathrm{cm}$ from the wire, and give its direction.

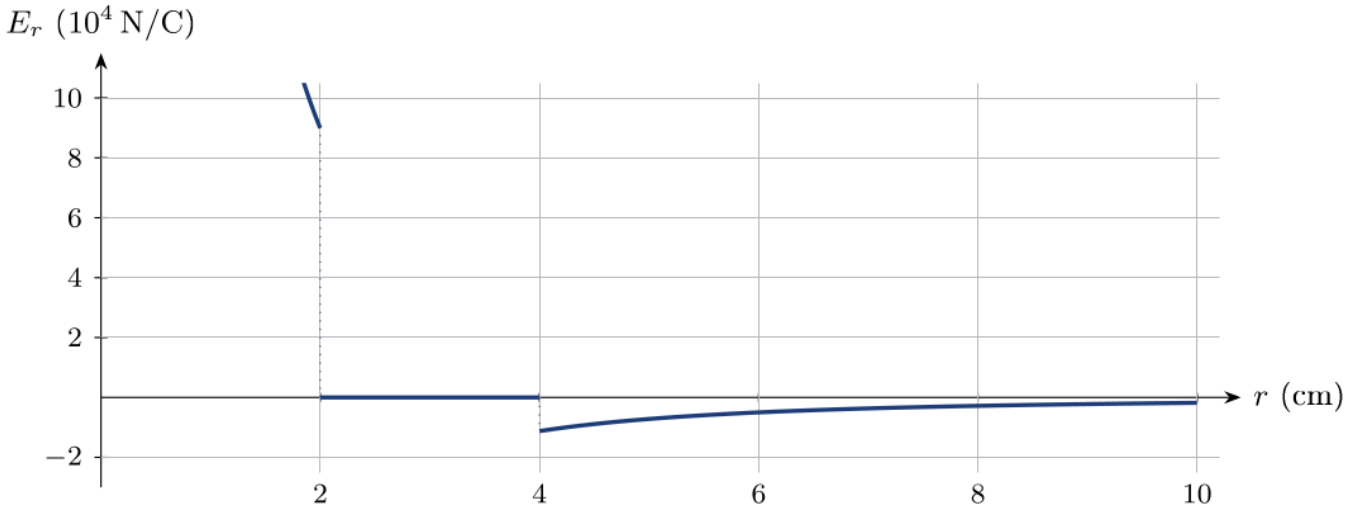
17.15 ★ A large, thin, nonconducting sheet carries $-3.0\ \mu\mathrm{C/m^2}$. Find the field $2.0\ \mathrm{mm}$ and $2.0\ \mathrm{m}$ from it, and give its direction.

17.16 ★★ A solid insulating sphere of radius $6.0\ \mathrm{cm}$ carries $+15\ \mathrm{nC}$ spread uniformly through its volume. Find the field at $r = 2.0\ \mathrm{cm}$, $6.0\ \mathrm{cm}$, and $12\ \mathrm{cm}$.

17.17 ★★ A long, thin-walled metal pipe of radius $2.0\ \mathrm{cm}$ carries $6.0\ \mathrm{nC}$ per meter of length. Find the field (a) $1.0\ \mathrm{cm}$ from the axis and (b) $4.0\ \mathrm{cm}$ from the axis.

17.18 ★★ *[Real-world: electronics.]* A coaxial cable consists of a long central wire carrying charge $+\lambda$ per unit length, surrounded by a thin cylindrical conducting sheath carrying $-\lambda$ per unit length. Use Gauss's law to find the field (a) between the wire and the sheath and (b) outside the sheath. Explain why the cable produces no field outside.

17.19 ★★ *[Graph-reading.]* A point charge sits at the center of a thick, concentric conducting spherical shell. The graph shows the radial component of the field, E_r, against the distance r from the center (positive means outward).



•Figure description
A graph of the radial electric field in units of ten thousand newtons per coulomb against distance r in centimeters from 0 to 10. Just inside r equals 2 centimeters the field is large and positive, falling steeply to about 9 at r equals 2; at smaller r the curve rises above the top of the graph. Between r equals 2 and 4 centimeters the field is exactly zero. Beyond 4 centimeters the field is negative, starting at about minus 1.1 and approaching zero as r increases, reaching about minus 0.3 at 8 centimeters.

a. What are the inner and outer radii of the conducting shell? How can you tell?
b. Find the point charge at the center.
c. Find the net charge on the shell, and how it is divided between its inner and outer surfaces.

17.20 ★★★ A long insulating cylinder of radius R carries a volume charge density that increases with distance from its axis, $\rho = \rho_0 r/R$. Find the field inside and outside the cylinder, and check that your two answers agree at $r = R$.

17.21 ★★★ A solid insulating sphere of radius a carries uniform charge density ρ. A spherical cavity of radius b is hollowed out of it, with the cavity's center a distance d from the sphere's center. Show that the field everywhere inside the cavity is uniform, $\mathbf{E} = \rho\mathbf{d}/3\epsilon_0$, where $\mathbf{d}$ points from the sphere's center to the cavity's center. *(Hint: the cavity is equivalent to the full sphere plus a small sphere of charge density $-\rho$. Write the field of each inside itself in vector form, $\mathbf{E} = \rho\mathbf{r}/3\epsilon_0$.)*

Section 17.5 — Conductors in Electrostatic Equilibrium

17.22 ★ A charged conductor has a surface charge density of $4.0\ \mu\text{C/m}^2$ at a certain point on its surface. Find the field just outside the conductor there, and just inside.

17.23 ★ *[Conceptual.]* Explain why the occupants of a metal car struck by lightning are protected, and why a convertible with a fabric roof would offer much less protection.

17.24 ★★ A solid conducting sphere of radius $8.0\ \text{cm}$ carries a net charge of $+12\ \text{nC}$.

a. Find the surface charge density.
b. Find the field just outside the surface, and check that it equals σ/ϵ_0.
c. What is the field at $r = 4.0\ \text{cm}$?

17.25 ★★ A large, flat metal plate of area $0.50\ \text{m}^2$ carries a total charge of $8.0\ \mu\text{C}$, spread evenly over its two faces. Find the field just outside each face. Compare with the field of a thin nonconducting sheet carrying the same total charge on the same area, and explain the agreement.

17.26 ★★ A point charge of $-3.0\ \text{nC}$ sits at the center of a neutral conducting spherical shell with inner radius $5.0\ \text{cm}$ and outer radius $7.0\ \text{cm}$.

a. Find the charge on the inner and outer surfaces of the shell, and the surface charge density on each.
b. Find the field at $r = 3.0\ \text{cm}$, $6.0\ \text{cm}$, and $10\ \text{cm}$.
c. The shell's outer surface is now grounded. Which of your answers change?

17.27 ★★★ *[Real-world: atmospheric physics.]* On a fair day the field is about $150\ \text{N/C}$ downward at the ground, but only about $100\ \text{N/C}$ downward at an altitude of $300\ \text{m}$. Treating the atmosphere over a small region as having planar symmetry, find the average charge density in the air between the ground and $300\ \text{m}$. Is it positive or negative?

17.28 ★★★ *[Real-world: atomic physics.]* In a simple model of the hydrogen atom, the proton is a point charge $+e$ at the center, surrounded by an electron cloud of total charge $-e$ with charge density $\rho(r) = -(e/\pi a_0^3)\ e^{-2r/a_0}$, where $a_0 = 52.9\ \text{pm}$ is the Bohr radius (Chapter 41).

a. Show that the total charge of the cloud is $-e$. *(You may use* $\int_0^\infty x^2 e^{-x}\ dx = 2$*.)*
b. Find the net charge enclosed by a sphere of radius a_0, and the field there. *(Use* $\int_0^2 x^2 e^{-x}\ dx = 2 - 10e^{-2}$*.)*
c. Explain why the atom's field falls to zero far away much faster than $1/r^2$.

17.29 ★★★ Two very long, coaxial, thin-walled cylindrical shells have radii a and $b > a$. The inner shell carries charge $+\lambda$ per unit length and the outer shell -2λ per unit length.

a. Find the field in all three regions.
b. Sketch E_r against r.
c. If the outer shell were a thick conductor rather than a thin shell, how would its charge divide between its inner and outer surfaces?

This page titled 17.E: Gauss's Law (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18: ELECTRIC POTENTIAL

A flashlight battery is stamped “1.5 V,” a car battery “12 V,” and a wall outlet in North America delivers about 120 volts. High-voltage transmission lines carry hundreds of thousands of volts across the countryside, and a thundercloud may sit a hundred million volts away from the ground below it. The volt is probably the most familiar electrical unit in everyday life, yet so far this book has not needed it. Coulomb’s law, electric fields, and Gauss’s law are all about *forces*. The volt is about *energy*.

In Part I, energy turned many hard problems into easy ones. Instead of following a roller coaster’s acceleration point by point, we compared its kinetic and potential energy at two positions, and the answer followed in a single line. The same shortcut works for electric forces, because like gravity they are conservative. This chapter develops the **electric potential**, the electric potential energy per unit charge. It is a scalar, which makes it far easier to add up than the field, and once it is known, the field itself follows from it by differentiation. Along the way we will find out why lightning rods are pointed, why a charge inside a hollow conductor is shielded from the outside world, and what a “300-volt electron” really means.

Learning Objectives

- Define the electric potential and potential difference, and relate them to the work done by the electric field.
- Use electric potential energy and energy conservation, including the electron volt, to find the speeds of charged particles.
- Sketch equipotential surfaces, relate them to field lines, and explain why a conductor is an equipotential.
- Calculate the potential of a point charge, an electric dipole, a system of point charges, and continuous charge distributions, and the potential energy of a system of charges.
- Find the electric field from the potential.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

18.1: Electric Potential
18.2: Equipotential Surfaces
18.3: Potential of a Discrete Charge Arrangement
18.4: Potential of a Continuous Charge Arrangement
18.5: Calculating the Field from the Potential
18.S: Electric Potential (Summary)
18.E: Electric Potential (Problems)

This page titled 18: Electric Potential is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.1: Electric Potential

In Chapter 8 a force was called **conservative** if the work it does on an object moving between two points does not depend on the path taken. Gravity is conservative, and that is what allowed us to define gravitational potential energy: a single number, U, attached to each position, whose change $\Delta U = -W$ records the work gravity does. The electric force between charges at rest is also conservative. Coulomb's law has exactly the form of Newton's law of gravitation, a force directed along the line between two point charges whose size depends only on their separation, and the argument that made gravity conservative applies without change. By superposition, the electric field of *any* arrangement of static charges is conservative as well.

Suppose a test charge q_0 moves from point A to point B in an electric field $\mathbf{E}$. The field exerts a force $q_0\mathbf{E}$ on it, and does work

$$W_{AB} = \int_A^B q_0 \mathbf{E}\cdot d\mathbf{s}, \tag{18.1.1}$$

where $d\mathbf{s}$ is a small step along the path. Because the force is conservative, W_{AB} depends only on the endpoints, and the change in the electric potential energy is $\Delta U = -W_{AB}$. That change is proportional to q_0. Dividing it out, just as we divided the force by q_0 to define the field, leaves a quantity that depends only on the field and the two points. It is the **potential difference** between B and A:

Key Equation: Potential difference

$$\Delta V = V_B - V_A = \frac{\Delta U}{q_0} = -\int_A^B \mathbf{E}\cdot d\mathbf{s}. \tag{18.1}$$

The **electric potential** V at a point is the electric potential energy per unit charge that a test charge would have there. Like potential energy, it is defined only up to a constant: only *differences* in potential have physical meaning, and we are free to choose where $V = 0$. For a charge distribution of limited size the natural choice is $V = 0$ infinitely far away. In circuits, the conventional zero is the Earth, or *ground*. The potential difference between two points is also called the **voltage** between them.

The SI unit of potential is the **volt** (V), named for Alessandro Volta, who built the first battery in 1800:

$$1\,\mathrm{V} = 1\,\mathrm{J/C}. \tag{18.1.2}$$

A potential difference of one volt means that one joule of work is needed to move one coulomb of charge between the two points. The volt also gives a second name to the unit of electric field, as Chapter 16 promised: $1\ \mathrm{N/C} = 1\ \mathrm{N\cdot m/(C\cdot m)} = 1\ \mathrm{J/(C\cdot m)} = 1\ \mathrm{V/m}$. From now on we will usually quote fields in volts per meter.

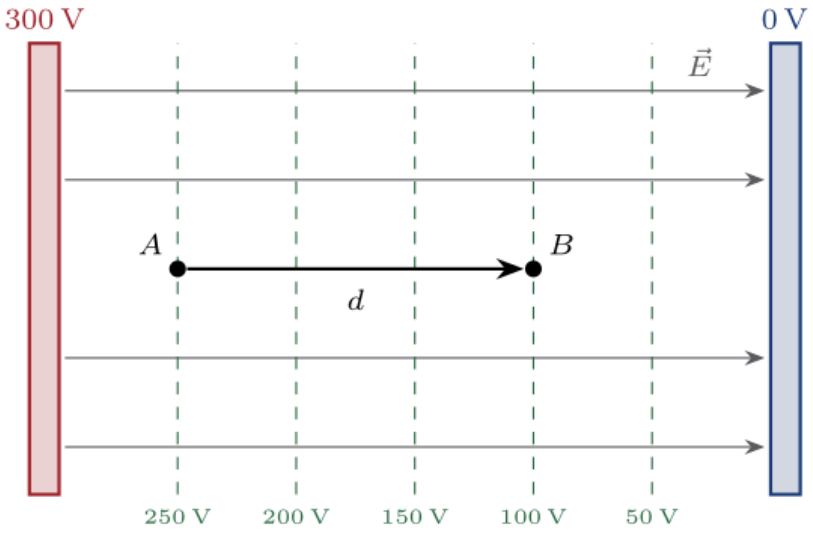


Figure 18.1. The potential in a uniform field. The field points from the positive plate toward the negative one, and the potential falls steadily in that direction. The dashed lines join points of equal potential and are perpendicular to $\mathbf{E}$. Moving from A to B, a distance d along the field, the potential falls by Ed.

•Figure description

Two large parallel plates, a positive plate on the left at 300 volts and a negative plate on the right at 0 volts, with a uniform electric field pointing from left to right between them. Dashed vertical lines between the plates mark positions of equal potential, labeled 250, 200, 150, 100 and 50 volts, decreasing from left to right. A point A near the left plate and a point B farther right are joined by a displacement along the field of length d.

A uniform field. The simplest case is a uniform field $\mathbf{E}$, such as the field between two large, oppositely charged parallel plates (Section 16.3). Move a test charge a displacement $\Delta\mathrm{s}$. Since $\mathbf{E}$ is constant it comes out of the integral, and

Key Equation: Potential difference in a uniform field

$$\Delta V = -\mathbf{E}\cdot\Delta\mathrm{s} = -Ed, \tag{18.2}$$

where d is the component of the displacement *along* the field. The minus sign carries the key physical idea: *the electric field points in the direction of decreasing potential*. Moving with the field, the potential drops; moving against it, the potential rises; moving perpendicular to it, the potential does not change at all (Figure 18.1).

Example 18.1: The potential difference between two plates

Two large parallel plates $1.50\ \mathrm{cm}$ apart have a uniform field of $2.00\times10^{4}\ \mathrm{V/m}$ between them. (a) What is the potential difference between the plates, and which plate is at the higher potential? (b) How much work does the field do on a charge of $+3.0\ \mathrm{nC}$ carried from the positive plate to the negative plate?

Solution

(a) The field points from the positive plate to the negative plate, so moving in that direction is moving along the field, and by equation (18.2)

$$\Delta V = V_- - V_+ = -Ed = -(2.00\times10^{4}\,\mathrm{V/m})(0.0150\,\mathrm{m}) = -300\,\mathrm{V}. \tag{18.1.3}$$

The potential difference between the plates is $\boxed{300\,\mathrm{V}}$, with the *positive* plate higher. (This is the same pair of plates that accelerated an electron in Example 16.10.)

(b) The change in potential energy is $\Delta U = q\ \Delta V = (3.0\times10^{-9})(-300) = -9.0\times10^{-7}$ J, and the work done by the field is $W = -\Delta U = \boxed{+9.0\times10^{-7}\,\mathrm{J}}$. The field pushes the positive charge "downhill" in potential and does positive work on it, exactly as gravity does positive work on a falling rock.

Example 18.2: The path does not matter

A uniform field of $500\ \mathrm{V/m}$ points in the $+x$ direction. Find $V_B - V_A$ between A at the origin and B at $(0.40\ \mathrm{m},\ 0.30\ \mathrm{m})$, (a) along the straight line from A to B, and (b) along the path that runs first along x to $C = (0.40\ \mathrm{m},\ 0)$ and then along y to B.

Solution

(a) The displacement is $\Delta\mathbf{s} = (0.40\ \hat{\imath} + 0.30\ \hat{\jmath})\ \mathrm{m}$, and $\mathbf{E} = 500\ \hat{\imath}\ \mathrm{V/m}$, so

$$V_B - V_A = -\mathbf{E}\cdot\Delta\mathbf{s} = -(500)(0.40) = \boxed{-200\ \mathrm{V}}. \tag{18.1.4}$$

(b) From A to C the step lies along the field, so $V_C - V_A = -(500)(0.40) = -200\ \mathrm{V}$. From C to B the step is perpendicular to the field, and $V_B - V_C = 0$. The total is again $\boxed{-200\ \mathrm{V}}$. Any path gives the same answer, because only the displacement along the field matters. That is what it means for the potential to be well defined.

Note: Potential versus potential energy

The two terms are easy to confuse. *Potential energy* U, in joules, belongs to a charge in a field; it depends on the charge. *Potential* V, in volts, belongs to a point in space; it is set by the source charges alone, whether or not a test charge is there. The relationship is exactly that between force and field: $\mathbf{F} = q\mathbf{E}$, and $U = qV$.

Checkpoint 18.1

In a certain region the electric field points due north. (a) If you walk due north, does the potential increase, decrease, or stay the same? (b) What if you walk due east? (c) A point P is at a potential of $50\ \mathrm{V}$ and a point Q is at $-20\ \mathrm{V}$. Is the field, on average, pointing from P toward Q or the other way?

Answer

Electric Potential Energy

Once the potential is known at every point, the **electric potential energy** of any charge q placed at a point follows at once, by turning the definition around:

Key Equation: Electric potential energy of a charge

$$U = qV, \qquad \Delta U = q\,\Delta V. \tag{18.3}$$

If the electric force is the only force doing work, mechanical energy is conserved, as in Chapter 8:

$$K_i + qV_i = K_f + qV_f, \qquad \text{or} \qquad \Delta K = -q\,\Delta V. \tag{18.4}$$

A charge that gains kinetic energy must lose potential energy. For a *positive* charge that means moving to a lower potential: positive charges "fall downhill" in potential, along the field. A *negative* charge loses potential energy by moving to a *higher* potential, so electrons "fall uphill," against the field. Keeping the sign of q in equation (18.4) takes care of this automatically.

The electron volt. Atomic and subatomic particles have such small energies that the joule is an awkward unit for them. A more convenient one is the energy a particle of charge e gains in falling through a potential difference of one volt. This is the **electron volt**:

$$1\ \mathrm{eV} = (1.602\times10^{-19}\ \mathrm{C})(1\ \mathrm{V}) = 1.602\times10^{-19}\ \mathrm{J}. \tag{18.5}$$

An electron or proton accelerated from rest through a potential difference of $300\ \mathrm{V}$ therefore gains exactly $300\ \mathrm{eV}$ of kinetic energy, without any arithmetic. Chemical reactions involve energies of a few eV per molecule, X-ray photons thousands (keV), and nuclear reactions millions (MeV). You will meet all of these in Part III.

Example 18.3: Accelerating an electron, by energy

An electron is released from rest at the negative plate of Example 18.1 and accelerates to the positive plate. Find its kinetic energy in electron volts and joules, and its final speed.

Solution

The electron moves from the negative plate to the positive plate, up $300\ \mathrm{V}$ in potential, so $\Delta V = +300\ \mathrm{V}$. By equation (18.4), with $q = -e$,

$$\Delta K = -q\,\Delta V = -(-e)(300\,\mathrm{V}) = \boxed{300\,\mathrm{eV}} = (300)(1.602\times10^{-19}\,\mathrm{J}) = 4.81\times10^{-17}\,\mathrm{J}. \tag{18.1.5}$$

Starting from rest, $K = \frac{1}{2}m_e v^2$, so

$$v = \sqrt{\frac{2K}{m_e}} = \sqrt{\frac{2(4.81\times10^{-17})}{9.11\times10^{-31}}} = \boxed{1.03\times10^{7}\,\mathrm{m/s}}. \tag{18.1.6}$$

This is exactly the answer Example 16.10 found by computing the acceleration, then using a kinematic equation. The energy method needs neither the acceleration nor the distance between the plates, only the potential difference. It works just as well when the field is not uniform, where the kinematic method would fail.

Example 18.4: Stopping a proton

A proton moving at $3.0\times10^{5}\ \mathrm{m/s}$ heads directly into a region where the potential rises. Through what potential difference must it pass to be brought to rest?

Solution

The proton's kinetic energy is $K = \frac{1}{2}(1.673\times10^{-27})(3.0\times10^{5})^2 = 7.5\times10^{-17}\ \mathrm{J}$, or $K/e = 470\ \mathrm{eV}$. It stops when all of this has become potential energy, $e\ \Delta V = K$, so

$$\Delta V = \frac{K}{e} = \boxed{470\,\mathrm{V}}. \tag{18.1.7}$$

Once K is in electron volts, the stopping potential for a singly charged particle is the same number in volts.

Checkpoint 18.2

A proton and an electron are each released from rest midway between two charged plates. (a) Which way does each move, toward higher or lower potential? (b) When each reaches a plate, which has gained more kinetic energy? (c) Which is moving faster?

Answer

Answers to Checkpoints

Checkpoint 18.1. (a) It decreases: the field points in the direction of *decreasing* potential. (b) It stays the same. Walking perpendicular to the field, $\mathbf{E}\cdot d\mathbf{s} = 0$, so you are moving along an equipotential. (c) From P toward Q: the field points from high potential ($50\ \mathrm{V}$) toward low potential ($-20\ \mathrm{V}$).

Checkpoint 18.2. (a) Each moves in the direction that lowers its potential energy $U = qV$. The proton, positive, moves toward *lower* potential (along the field). The electron, negative, moves toward *higher* potential (against the field). (b) They gain the same kinetic energy. Each starts midway and ends at a plate, so each falls through a potential difference of the same size, half the plate voltage, and $|q| = e$ for both. The kinetic energy gained, $e\ |\Delta V|$, is the same. (c) The electron: with equal kinetic energies, $v \propto 1/\sqrt{m}$, so the electron moves about $\sqrt{1840} \approx 43$ times faster.

This page titled 18.1: Electric Potential is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.2: Equipotential Surfaces

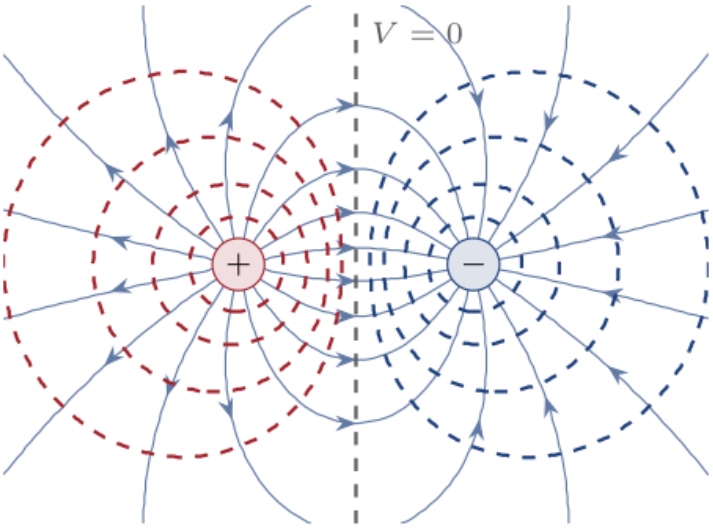


Figure 18.2. Equipotentials (dashed) and field lines (light blue) of an electric dipole, computed from the exact potential. The equipotentials are drawn at equal steps of potential, red for positive and blue for negative, and the plane midway between the charges is the equipotential $V = 0$. Field lines and equipotentials cross at right angles everywhere.

•Figure description

Computed equipotentials and field lines of an electric dipole, a positive charge on the left and an equal negative charge on the right. The equipotentials are closed curves: red loops of positive potential surround the positive charge, blue loops of negative potential surround the negative charge, both shrinking toward the charges. A straight gray vertical line midway between the charges is the zero-potential equipotential. The field lines, drawn as thin blue lines, curve from the positive to the negative charge and cross every equipotential at right angles.

A surface on which the potential has the same value everywhere is called an **equipotential surface**, or simply an equipotential. The dashed lines in Figure 18.1 are equipotentials seen edge-on. Equipotentials are to the potential what contour lines on a topographic map are to height, and they have three important properties.

1. *No work is done moving a charge along an equipotential.* Since $\Delta V = 0$, equation (18.3) gives $\Delta U = 0$.
2. *The field is perpendicular to the equipotentials everywhere.* If $\mathbf{E}$ had a component along an equipotential surface, a small step along the surface would change the potential by $-\mathbf{E} \cdot d\mathbf{s} \neq 0$, contradicting property 1. The field points from higher equipotentials toward lower ones, the direction of steepest descent.
3. *Equipotentials drawn at equal steps of potential crowd together where the field is strong.* For a small step Δs perpendicular to the equipotentials, $|\Delta V| \approx E\ \Delta s$, so a fixed ΔV is crossed in a shorter distance where E is larger, just as closely spaced contour lines on a map mark a steep hillside.

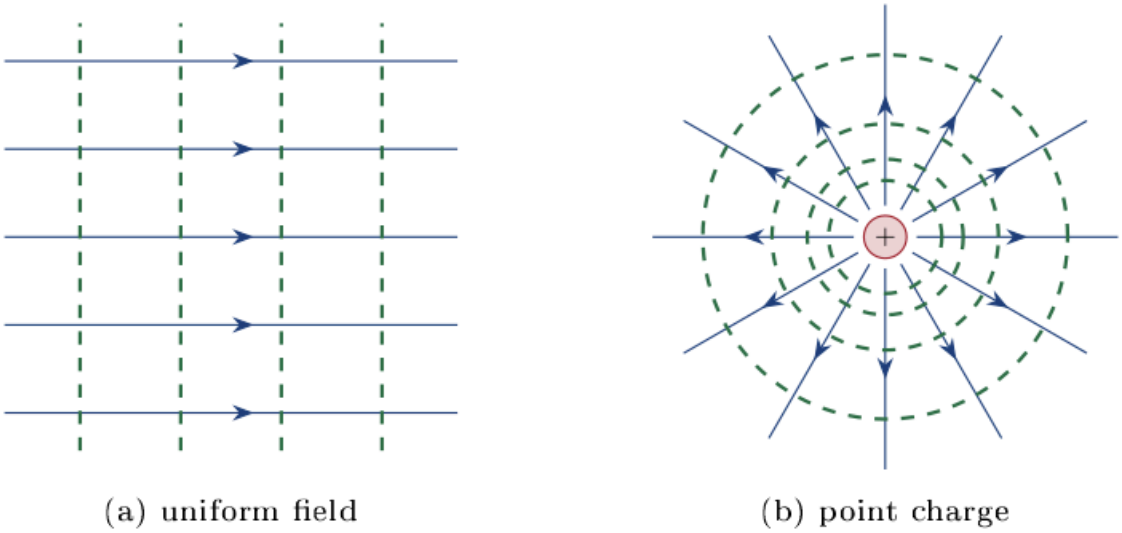


Figure 18.3. Equipotentials (dashed green) and field lines (solid blue). (a) A uniform field has flat, evenly spaced equipotentials. (b) A point charge has spherical equipotentials. Drawn at equal steps of potential, they crowd together near the charge, where the field is strongest. In both cases the field lines cross the equipotentials at right angles.

•Figure description

Two panels of equipotentials, drawn dashed, and field lines, drawn solid with arrows. In panel a, a uniform field is shown as parallel horizontal field lines pointing right, crossed at right angles by evenly spaced vertical equipotential lines. In panel b, a positive point charge sits at the center of radial field lines pointing outward, crossed at right angles by concentric circular equipotentials that are close together near the charge and farther apart away from it.

Figure 18.3 shows the equipotentials of a uniform field, which are parallel planes, and of a point charge, which are concentric spheres, crowded near the charge where the field is strong. Figure 18.2 shows those of an electric dipole, together with its field lines. Everywhere the two families cross at right angles.

Conductors are equipotentials. Inside a conductor in electrostatic equilibrium the field is zero (Section 17.5). By equation (18.1), the potential difference between any two points in the metal is then zero: *every point of a conductor in equilibrium, inside and on its surface, is at the same potential*. Its surface is an equipotential, which is another way of seeing that the field just outside it must be perpendicular to the surface.

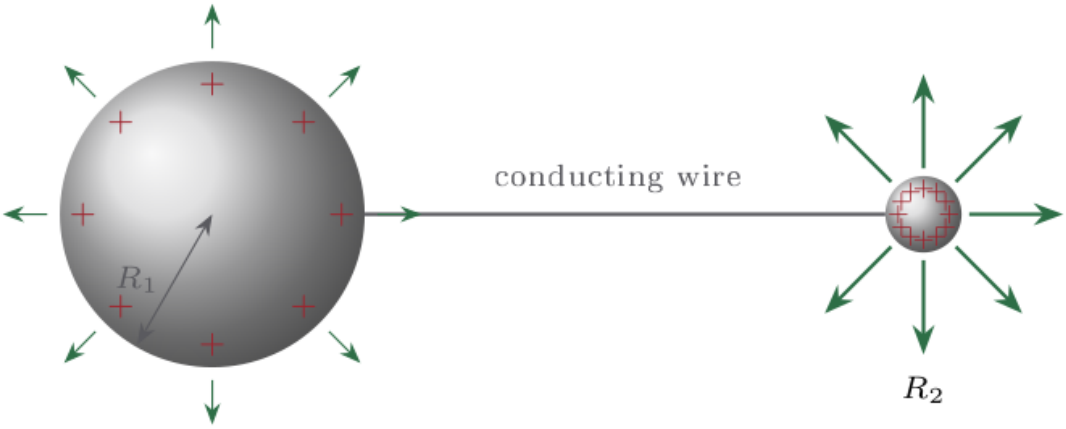


Figure 18.4. Two conducting spheres joined by a wire are at the same potential, so $Q_1/R_1 = Q_2/R_2$ and the surface charge density satisfies $\sigma \propto 1/R$. The smaller sphere has the denser charge and the stronger field just outside it (green arrows), which is why charge and field concentrate at sharp points.

•Figure description

Two conducting spheres joined by a long thin wire: a large sphere of radius R sub 1 on the left and a small sphere of radius R sub 2 on the right. Both carry positive charge; the plus signs on the small sphere are packed much more closely than those on the large sphere, showing its higher surface charge density. Short arrows showing the field just outside each surface are longer on the small sphere.

Why charge crowds at sharp points. Chapter 17 claimed that on an irregular conductor the charge density is greatest where the surface curves most sharply. The equipotential property explains why. Consider two conducting spheres of radii R_1 and R_2, far apart, joined by a long thin wire so that they form a single conductor (Figure 18.4). The whole arrangement is one equipotential, so the two spheres have the same potential. The potential of a charged sphere is k_eQ/R (Section 18.3), so

$$\frac{k_eQ_1}{R_1} = \frac{k_eQ_2}{R_2} \qquad \Longrightarrow \qquad \frac{Q_1}{R_1} = \frac{Q_2}{R_2}. \tag{18.2.1}$$

The larger sphere holds more charge, but in proportion only to its radius, while its area grows as the *square* of its radius. Writing $Q = 4\pi R^2\sigma$ gives $\sigma_1 R_1 = \sigma_2 R_2$: the surface charge density is inversely proportional to the radius. The field just outside is σ/ϵ_0 (equation (17.6)), so it too is largest on the smaller sphere. A sharp point on a conductor behaves like a very small sphere, and the field there can be strong enough to ionize the air. This is the corona discharge of Section 17.5, and it is why a lightning rod is pointed: its tip leaks charge into the air gradually and offers the lightning a preferred path to ground.

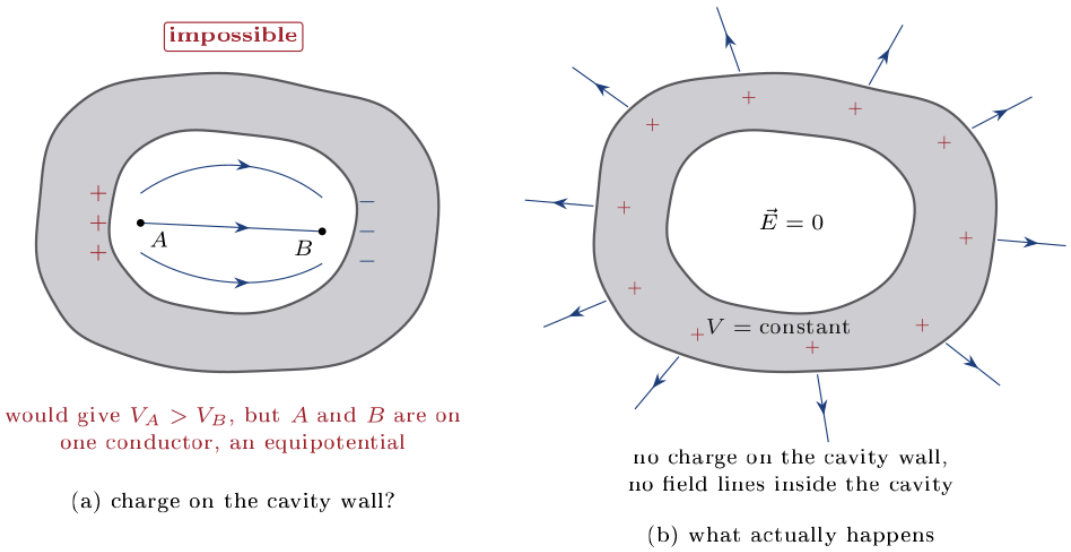


Figure 18.5. An empty cavity in a conductor. (a) If the cavity wall carried a positive patch and a negative patch, field lines would cross the cavity from one to the other, and the potential would drop along them from A to B. But the whole conductor is an equipotential, so $V_A = V_B$, and this arrangement is impossible. (b) Any charge on the conductor sits on its outer surface. The cavity wall carries no charge, and the field everywhere inside the cavity is zero, whatever the charges and fields outside.

•Figure description

Two panels, each showing a solid conductor, shaded gray, with an empty cavity, shown white, inside it. Panel a, labeled impossible: a patch of positive charge sits on the left side of the cavity wall and a patch of negative charge on the right side. Field lines cross the empty cavity from the positive patch to the negative patch. Following a field line from point A on the positive patch to point B on the negative patch, the potential drops, so V sub A would be greater than V sub B. A note says this cannot happen, because A and B are on the same conductor, which is an equipotential. Panel b, what actually happens: positive charge sits only on the outer surface of the conductor, with field lines pointing outward from it into the space outside. The cavity wall carries no charge, there are no field lines in the cavity, and the cavity is labeled E equals zero. The whole conductor, including the cavity, is labeled V equals constant.

Example 18.5: Two connected spheres

Two conducting spheres, of radii $R_1 = 10.0\ \text{cm}$ and $R_2 = 2.0\ \text{cm}$, are far apart but joined by a thin wire, and together carry $+60\ \text{nC}$. Find the charge on each sphere and the field just outside each surface.

Solution

Equal potentials require $Q_1/R_1 = Q_2/R_2$, so $Q_1 = 5Q_2$. With $Q_1 + Q_2 = 60\ \text{nC}$, the charges are $Q_2 = \boxed{10\,\text{nC}}$ and $Q_1 = \boxed{50\,\text{nC}}$. The field just outside each sphere is $k_e Q/R^2$:

$$E_1 = \frac{(8.99\times10^9)(50\times10^{-9})}{(0.100)^2} = \boxed{4.5\times10^4\,\text{V/m}},$$
$$E_2 = \frac{(8.99\times10^9)(10\times10^{-9})}{(0.020)^2} = \boxed{2.2\times10^5\,\text{V/m}}. \tag{18.2.2}$$

The small sphere carries a fifth as much charge but has five times the surface field, in keeping with $E \propto \sigma \propto 1/R$.

An empty cavity in a conductor. Chapter 17 showed that the walls of an empty cavity in a conductor carry no *net* charge, and promised that they carry no charge at all. Suppose the wall had a patch of positive charge somewhere and a patch of negative charge elsewhere, adding to zero (Figure 18.5a). Then field lines would leave the positive patch, cross the empty cavity, and end on the negative patch. Integrating $-\mathbf{E}\cdot d\mathbf{s}$ along one of those field lines gives a potential *drop* from the first patch to the second, since the field points along the path all the way. But both patches are on the same conductor, which is an equipotential. The contradiction means there can be no such field lines: the cavity wall carries no charge anywhere, and the field inside the cavity is zero (Figure 18.5b). This is the full justification for the Faraday cage of Section 17.5.

Checkpoint 18.3

(a) Can two different equipotential surfaces ever intersect? Explain. (b) A charged conductor has a flat face and a pointed tip. Where is the potential higher, on the face or at the tip? Where is the field just outside higher?

Answer

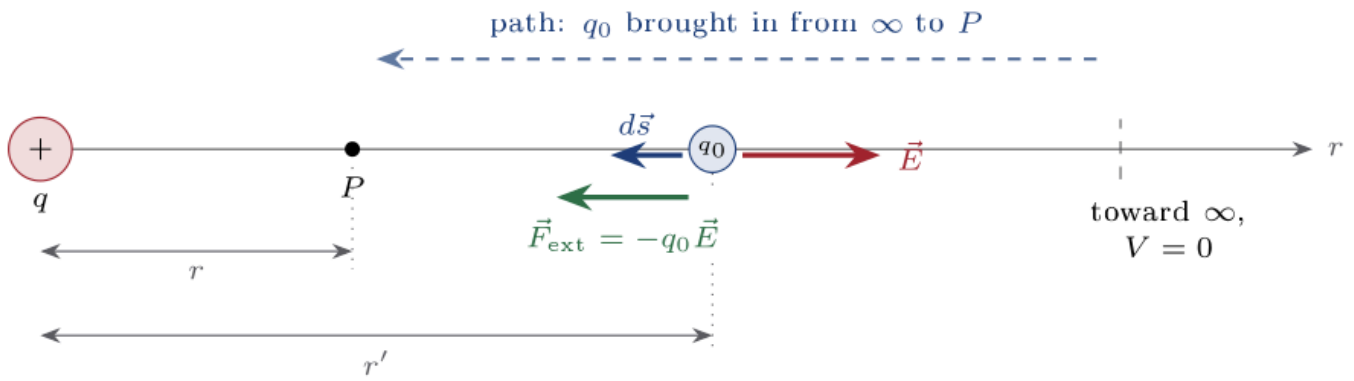
Answers to Checkpoints

Checkpoint 18.3. (a) No. Every point has a single value of potential. At an intersection that point would have to be at two different potentials at once. (b) The potential is the same everywhere on the conductor (face and tip alike) because a conductor in equilibrium is an equipotential. The field just outside is much stronger at the tip, because charge crowds onto sharply curved regions, and $E = \sigma/\epsilon_0$ is larger where σ is larger.

This page titled 18.2: Equipotential Surfaces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.3: Potential of a Discrete Charge Arrangement

We now calculate the potential produced by point charges. One point charge is enough to give the basic result. The rest follows from superposition, and here the potential shows its great practical advantage: it is a scalar. The potentials of several charges simply add, with no directions or components to keep track of.



$$V(r) - V(\infty) = -\int_{\infty}^{r} \vec{E} \cdot d\vec{s} = \frac{k_e q}{r}$$

Figure 18.6. Finding the potential of a point charge. A test charge q_0 is brought in along a radial line from infinity, where $V = 0$, to the point P at distance r. At each step $d\mathbf{s}$ points inward while $\mathbf{E}$ points outward, so $\mathbf{E} \cdot d\mathbf{s}$ is negative, and the potential rises as the charge comes in. An outside agent must push with force $-q_0\mathbf{E}$ to move the charge slowly, and the work it does, per unit charge, is the potential at P.

•Figure description

A positive point charge q sits at the left end of a horizontal radial axis labeled r, which runs off to the right toward infinity, where the potential is zero. A dashed arrow along the axis shows the path of a test charge q sub zero being brought in from infinity, at the far right, to a point P a distance r from q. Partway along the path, at distance r prime, the test charge is drawn with three arrows: the field E pointing outward, away from q; the small step d s pointing inward, toward q; and the force of the outside agent pushing it inward, equal to minus q sub zero times E. Below the axis, the result is written: the potential at P minus the potential at infinity equals minus the integral from infinity to r of E dot d s, which equals k sub e q over r.

Potential of a Charged Particle

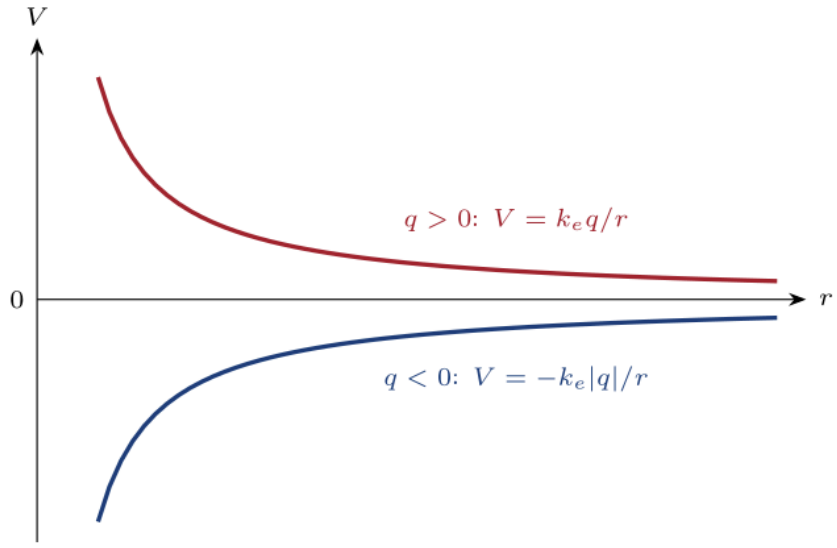


Figure 18.7. The potential of a point charge. It is positive near a positive charge and negative near a negative charge, and in both cases it falls toward zero as $1/r$.

•Figure description

A graph of electric potential V against distance r from a point charge. For a positive charge the curve, drawn in red, is positive and falls from large values near r equals zero toward zero as r increases, like one over r. For a negative charge the curve, drawn in blue, is its mirror image below the axis, rising from large negative values toward zero.

Take a point charge q at the origin and set $V = 0$ infinitely far away. To find the potential at distance r, imagine bringing a test charge q_0 in from infinity, where $V = 0$, to the point P at distance r, and integrate equation (18.1) along that radial path (Figure 18.6). At a distance r' along the way, $\mathbf{E} \cdot d\mathbf{s} = E\ dr' = (k_e q/r'^2)\ dr'$, where dr' is negative because the step points inward:

$$V(r) - V(\infty) = -\int_{\infty}^{r} \frac{k_e q}{r'^2}\, dr' = \left[\frac{k_e q}{r'}\right]_{\infty}^{r} = \frac{k_e q}{r}. \tag{18.3.1}$$

Any path would give the same result. Equivalently, $V(r)$ is the work per unit charge that an outside agent must do to bring a test charge slowly in from infinity to P. The **potential of a point charge** is

Key Equation: Potential of a point charge

$$V = \frac{k_e q}{r} = \frac{q}{4\pi\epsilon_0 r}. \tag{18.6}$$

The potential carries the sign of the charge: it is positive near a positive charge and negative near a negative one, and in both cases it approaches zero far away (Figure 18.7). Notice that it falls off as $1/r$, not $1/r^2$. By the shell theorem (Section 17.4), equation (18.6) also gives the potential outside any spherically symmetric distribution of total charge q, which is the result $V = k_e Q/R$ used above for a charged conducting sphere.

Example 18.6: The potential in a hydrogen atom

(a) Find the potential due to the proton at the electron's orbit in hydrogen, $r = 5.29 \times 10^{-11}$ m. (b) What is the electron's potential energy there, in electron volts?

Solution

(a) From equation (18.6),

$$V = \frac{k_e e}{r} = \frac{(8.99 \times 10^9)(1.602 \times 10^{-19})}{5.29 \times 10^{-11}} = \boxed{27.2\,\text{V}}. \tag{18.3.2}$$

(b) With $q = -e$, $U = qV = (-e)(27.2\ \text{V}) = \boxed{-27.2\,\text{eV}}$. The negative sign means the electron is bound: energy must be supplied to pull it away to infinity, where $U = 0$. Chapter 41 shows that the electron's kinetic energy in this orbit is $13.6\ \text{eV}$, so its total energy is $-13.6\ \text{eV}$, and $13.6\ \text{eV}$ is exactly the energy needed to ionize a hydrogen atom.

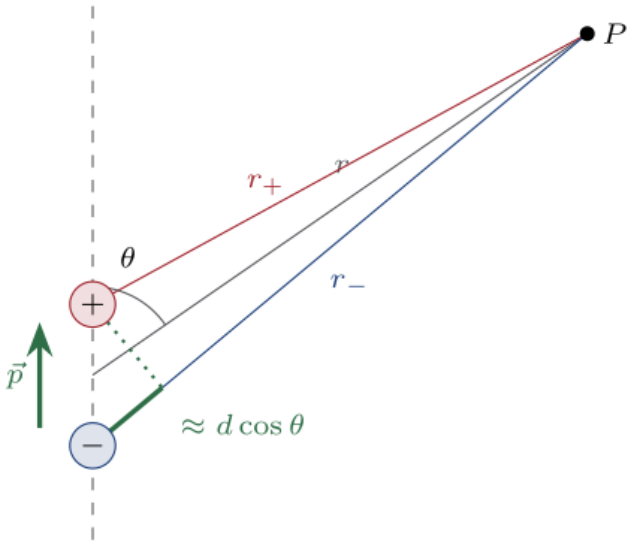


Figure 18.8. Geometry for the potential of a dipole. Far away the lines to P from the two charges are nearly parallel, and the negative charge is farther from P by about $d\cos\theta$.

•Figure description

An electric dipole along a vertical axis, with the positive charge above and the negative charge below, separated by d, and dipole moment p pointing up. A distant point P lies up and to the right, at distance r from the center, along a line making angle theta with the dipole axis. Lines from each charge to P are labeled r plus and r minus; they are nearly parallel. A short segment shows that r minus exceeds r plus by approximately d cosine theta.

Potential of an Electric Dipole

An electric dipole has charges $+q$ and $-q$ a distance d apart, with dipole moment $p = qd$ pointing from $-q$ to $+q$ (Section 16.2). At a point P a distance r_+ from the positive charge and r_- from the negative one (Figure 18.8), the two potentials simply add:

$$V = \frac{k_e q}{r_+} - \frac{k_e q}{r_-} = k_e q \frac{r_- - r_+}{r_+ r_-}. \tag{18.3.3}$$

Far from the dipole, $r \gg d$, the lines from the two charges to P are nearly parallel. The difference in their lengths is then $r_- - r_+ \approx d\cos\theta$, where θ is the angle between $\mathbf{p}$ and the line from the dipole's center to P, and the product $r_+ r_- \approx r^2$. So the **potential of an electric dipole** is

Key Equation: Potential of an electric dipole (far away)

$$V = \frac{k_e p \cos\theta}{r^2}. \tag{18.7}$$

Like the dipole's field, its potential depends only on p and falls off one power of r faster than a point charge's. It is positive on the side of the positive charge ($\theta < 90^\circ$), negative on the side of the negative charge, and zero everywhere on the perpendicular bisector ($\theta = 90^\circ$), where the two charges are equally far away. That the potential is zero there even though the field is not (Section 16.2) is a useful reminder that V and $\mathbf{E}$ measure different things: zero potential means zero work to bring a charge in from infinity, not zero force.

Example 18.7: The potential of a water molecule

Find the potential $1.0\ \text{nm}$ from a water molecule ($p = 6.2 \times 10^{-30}\ \text{C} \cdot \text{m}$), (a) along its axis on the positive side and (b) at 60° from its axis.

Solution

(a) On the axis $\theta = 0$, and

$$V = \frac{k_e p}{r^2} = \frac{(8.99 \times 10^9)(6.2 \times 10^{-30})}{(1.0 \times 10^{-9})^2} = \boxed{0.056\ \text{V}}. \tag{18.3.4}$$

(b) At 60°, $\cos\theta = 0.50$ and the potential is half as large, $\boxed{0.028\ \text{V}}$. A single elementary charge at the same distance would produce $k_e e/r = 1.4\ \text{V}$, some 25 times more.

Potential of a System of Charged Particles

For any number of point charges, the potential at a point is the sum of the potentials of the individual charges, each computed from equation (18.6):

Key Equation: Potential of a system of point charges

$$V = \sum_i \frac{k_e q_i}{r_i}, \tag{18.8}$$

where r_i is the distance from charge q_i to the point. This is an *algebraic* sum of positive and negative numbers. Compare the vector bookkeeping of Example 16.3.

Example 18.8: The potential near two charges

A charge $q_1 = +2.0\ \mu\text{C}$ sits at the origin and $q_2 = -3.0\ \mu\text{C}$ sits at $(0.40\ \text{m},\ 0)$. (a) Find the potential at $P = (0,\ 0.30\ \text{m})$. (b) How much work must be done by an outside agent to bring a charge of $+5.0\ \text{nC}$ in from very far away and place it at P?

Solution

(a) P is $0.30\ \text{m}$ from q_1 and $0.50\ \text{m}$ from q_2. From equation (18.8),

$$V_P = k_e\left(\frac{q_1}{r_1} + \frac{q_2}{r_2}\right) = (8.99 \times 10^9)\left(\frac{2.0 \times 10^{-6}}{0.30} - \frac{3.0 \times 10^{-6}}{0.50}\right) = \boxed{6.0 \times 10^3\ \text{V}}. \tag{18.3.5}$$

No components, no angles: just two numbers added with their signs.

(b) The charge starts where $V = 0$ and ends where $V = V_P$, at rest both times, so the work done by the outside agent equals the gain in potential energy:

$$W = q\,V_P = (5.0 \times 10^{-9})(6.0 \times 10^3) = \boxed{3.0 \times 10^{-5}\ \text{J}}. \tag{18.3.6}$$

The potential energy of a system of charges. Place a charge q_2 a distance r from a charge q_1. By equation (18.3), its potential energy in the potential of q_1 is $q_2(k_e q_1/r)$. This energy belongs to the *pair*, like the gravitational potential energy of Chapter 13:

Key Equation: Potential energy of two point charges

$$U = \frac{k_e q_1 q_2}{r}. \tag{18.9}$$

It equals the work needed to bring the two charges together from infinite separation. For like charges $U > 0$: work must be done

to push them together, and they fly apart if released. For unlike charges $U < 0$: they attract, and energy must be supplied to pull them apart. For three or more charges, the **potential energy of the system** is the sum of equation (18.9) over every *pair*, each pair counted once.

Example 18.9: Assembling three charges

How much work is needed to assemble three charges of $+1.0\ \mu\text{C}$ each at the corners of an equilateral triangle with sides of $10\ \text{cm}$, starting with the charges very far apart? If the charges are then released, what total kinetic energy do they eventually have?

Solution

There are three pairs, each separated by $a = 0.10\ \text{m}$:

$$U = 3\,\frac{k_e q^2}{a} = 3\,\frac{(8.99\times10^9)(1.0\times10^{-6})^2}{0.10} = \boxed{0.27\,\text{J}}. \tag{18.3.7}$$

That is the work required. When released, the charges repel and fly apart, and once they are far from one another their potential energy has all become kinetic energy: $\boxed{0.27\,\text{J}}$ in total.

Checkpoint 18.4

(a) Four equal positive charges sit at the corners of a square. Is the potential at the center zero? Is the field at the center zero? (b) Now make two adjacent corners positive and the other two negative, all of equal magnitude. Answer the same two questions.

Answer

Answers to Checkpoints

Checkpoint 18.4. (a) Not zero: potential is a scalar and the four positive contributions add, giving $V = 4k_e q/r$. The field is zero: the four field vectors cancel in opposite pairs. (b) Now the potential at the center is zero, since two $+k_e q/r$ terms cancel two $-k_e q/r$ terms. The field is *not* zero: the two positive charges push outward from their side and the two negative charges pull toward theirs, so all four contributions add up to a field pointing from the positive side toward the negative side. Potential and field answer different questions: one adds scalars, the other adds vectors.

This page titled 18.3: Potential of a Discrete Charge Arrangement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.4: Potential of a Continuous Charge Arrangement

Chapters 15 and 16 handled continuous distributions with a single idea: chop the charge into small elements dq, treat each element as a point charge, and add up the contributions with an integral. The potential works exactly the same way. By equation (18.6), an element dq a distance r from the point P contributes a potential

$$dV = k_e \frac{dq}{r}, \tag{18.4.1}$$

taking $V = 0$ at infinity. Potentials add, so the total is the integral of all the contributions:

Key Equation: Potential of a continuous charge distribution

$$V = \int dV = k_e \int \frac{dq}{r}. \tag{18.10}$$

Here r is the distance from dq to P. Set this beside the integral for the field, equation (16.8):

$$d\mathbf{E} = k_e \frac{dq}{r^2}\,\hat{r} \implies \mathbf{E} = \int d\mathbf{E}, \qquad dV = k_e \frac{dq}{r} \implies V = \int dV. \tag{18.4.2}$$

The recipe is the same, with two simplifications: the distance appears to the first power rather than the second, and there is no unit vector. Because dV is a scalar, there are no components to resolve and no symmetry argument is needed to cancel anything. The five steps of Section 15.4 carry over with their hardest step removed:

1. Sketch the distribution and the point P, and pick a typical element dq.
2. Write dq in terms of the charge density and a coordinate: $dq = \lambda\, dx$, $\sigma\, dA$, or $\rho\, dV$.
3. Write the distance r from dq to P in terms of that coordinate.
4. Integrate $dV = k_e\, dq/r$ over the whole distribution.
5. Check a limit, usually far away, where the result must become k_eQ/r.

A charged ring. A uniform ring of radius a carries charge Q. We want the potential at a point on its axis, a distance x from the center (see the drawing in Example 15.8). Every element dq of the ring is the same distance $r = \sqrt{x^2 + a^2}$ from that point, so r comes straight out of the integral, and $\int dq = Q$:

$$V_{\text{ring}} = \int k_e \frac{dq}{r} = \frac{k_e}{\sqrt{x^2 + a^2}} \int dq = \frac{k_e Q}{\sqrt{x^2 + a^2}}. \tag{18.11}$$

Far away, $x \gg a$, this becomes k_eQ/x, as it must. Compare the effort that the field, equation (16.11), took: there we needed the axial component $dE\cos\theta$ of each element's field and a symmetry argument to discard the rest.

A charged rod. A thin rod of length L carries charge Q uniformly, with $\lambda = Q/L$. Find the potential at a point P a distance y from the rod, on its perpendicular bisector. This is the same geometry as Figure 16.6. Put the rod on the x-axis from $-L/2$ to $L/2$. The element at x carries $dq = \lambda\, dx$ and is a distance $r = \sqrt{x^2 + y^2}$ from P, so

$$V = \int dV = k_e\lambda \int_{-L/2}^{L/2} \frac{dx}{\sqrt{x^2 + y^2}} = k_e\lambda \left[\ln\left(x + \sqrt{x^2 + y^2}\right)\right]_{-L/2}^{L/2}, \tag{18.4.3}$$

which simplifies to

$$V_{\text{rod}} = 2k_e\lambda\, \ln\left(\frac{L/2 + \sqrt{L^2/4 + y^2}}{y}\right). \tag{18.12}$$

Far away, $y \gg L$, the fraction is close to $1 + L/2y$, and since $\ln(1 + u) \approx u$ for small u, $V \approx 2k_e\lambda(L/2y) = k_eQ/y$, the potential of a point charge. Compare the field of the same rod in Section 16.3: there we needed both components of $d\mathbf{E}$, and a symmetry argument to cancel the components along the rod, before integrating. Here a single integral did everything.

Equation (18.12) also shows what goes wrong for an *infinite* line. For a long rod, $L \gg y$, the fraction is about L/y, so $V \approx 2k_e\lambda \ln(L/y)$, which grows without limit as $L \to \infty$. An infinite line has no finite potential relative to infinity. We return to this below.

A charged disk. Build a uniform disk of radius R and surface density σ from rings of radius r', width dr', and charge $dq = \sigma\ dA = 2\pi\sigma r'\ dr'$, as in Section 16.3. By equation (18.11), each ring contributes $dV = k_e\ dq/\sqrt{x^2 + r'^2}$ at a point on the axis, so

$$V = \int dV = \int_0^R \frac{2\pi k_e \sigma r'\, dr'}{\sqrt{x^2 + r'^2}} = 2\pi k_e \sigma \left[\sqrt{x^2 + r'^2}\right]_0^R, \tag{18.4.4}$$

which, using $2\pi k_e = 1/2\epsilon_0$, is

$$V_{\text{disk}} = \frac{\sigma}{2\epsilon_0}\left(\sqrt{x^2 + R^2} - x\right). \tag{18.13}$$

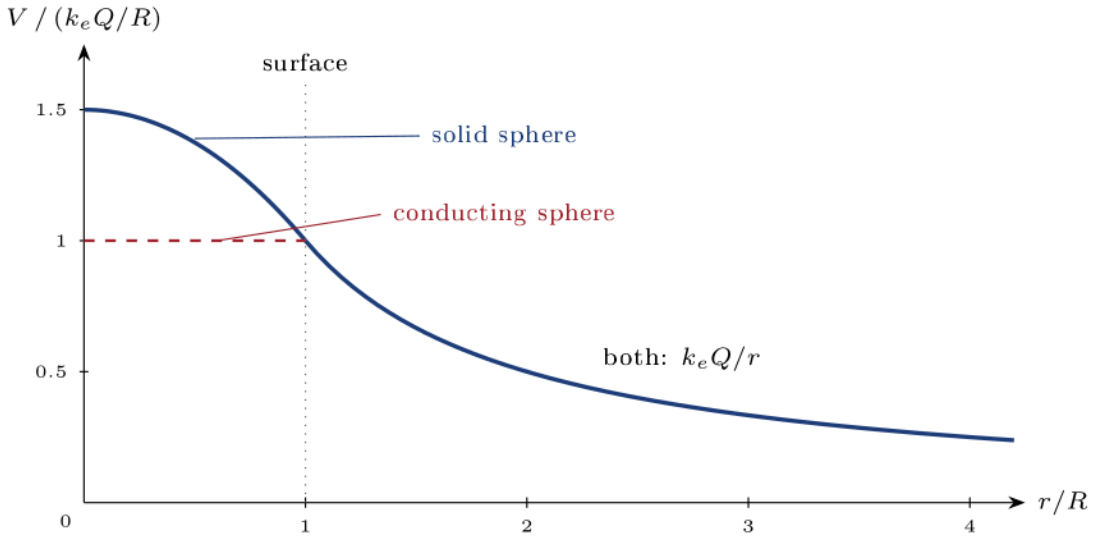


Figure 18.9. The potential of a charged sphere. Outside, both a uniformly charged solid sphere (solid) and a conducting sphere or shell (dashed) have the potential of a point charge. Inside, the conductor's potential stays at its surface value, while the solid sphere's rises to $\frac{3}{2}k_eQ/R$ at the center.

•Figure description

A graph of potential against distance r from the center of a charged sphere of radius R, for r from 0 to 4 R. Outside the sphere both curves are the same one over r curve, falling from k Q over R at the surface. Inside, the solid blue curve for a uniformly charged solid sphere rises smoothly to a rounded maximum of one and a half k Q over R at the center. The dashed red curve for a conducting sphere or thin shell stays flat at k Q over R everywhere inside.

Using the field instead. Adding up dV is not the only route. When the field is already known, from Gauss's law say, it is often easier to find the potential by integrating the field with equation (18.1). For a uniformly charged solid sphere of radius R and charge Q, the field outside is k_eQ/r^2, exactly as for a point charge, so outside the potential is also that of a point charge, $V = k_eQ/r$. At the surface it is k_eQ/R. Inside, the field is k_eQr/R^3 (equation (16.15)), and integrating inward from the surface,

$$V(r) = \frac{k_eQ}{R} - \int_R^r \frac{k_eQr'}{R^3}\, dr' = \frac{k_eQ}{R} + \frac{k_eQ}{2R^3}\left(R^2 - r^2\right), \tag{18.4.5}$$

which simplifies to

$$V = \frac{k_eQ}{2R}\left(3 - \frac{r^2}{R^2}\right) \quad (r \le R), \qquad V = \frac{k_eQ}{r} \quad (r \ge R). \tag{18.14}$$

The potential is greatest at the center, $\frac{3}{2}k_eQ/R$. For a charged *conducting* sphere, or a thin shell, the field inside is zero, so the potential inside does not change from its surface value: $V = k_eQ/R$ everywhere inside (Figure 18.9).

We can check the solid sphere's central value, $\frac{3}{2}k_eQ/R$, with $V = \int dV$ directly. Divide the ball into thin spherical shells of radius

r' and thickness dr'. A shell has volume $4\pi r'^2\ dr'$ and charge $dq = \rho\ 4\pi r'^2\ dr'$, where $\rho = 3Q/4\pi R^3$, and *every* part of it is the same distance r' from the center. So at the center

$$V = \int k_e \frac{dq}{r'} = \int_0^R 4\pi k_e \rho\, r'\, dr' = 2\pi k_e \rho R^2 = \frac{3}{2}\,\frac{k_e Q}{R}, \tag{18.4.6}$$

the same answer. Both routes are always available: add up $dV = k_e\ dq/r$, or integrate a known field. Choose whichever makes the integral easier.

An infinite line. Not every distribution allows $V = 0$ at infinity. An infinite line of charge extends to infinity itself. Summing dV over it diverges, as the long-rod limit of equation (18.12) showed, and so does integrating its field $\lambda/2\pi\epsilon_0 r$ in from infinity. We can still find the potential *difference* between two distances r_a and r_b from the line:

$$V_b - V_a = -\int_{r_a}^{r_b} \frac{\lambda}{2\pi\epsilon_0 r}\, dr = \frac{\lambda}{2\pi\epsilon_0} \ln\frac{r_a}{r_b}. \tag{18.15}$$

The same result follows from equation (18.12): for a long rod, $V_b - V_a \approx 2k_e\lambda\big[\ln(L/r_b) - \ln(L/r_a)\big] = 2k_e\lambda\ln(r_a/r_b)$, and the length L cancels. Only differences are physical, so this is all we ever need, and it is exactly what is needed for coaxial cables and the cylindrical capacitors of Chapter 19.

Example 18.10: Potentials of a ring, a disk, and a ball

Find the potential (a) on the axis of a ring of radius $3.0\ \text{cm}$ carrying $+12\ \text{nC}$, $4.0\ \text{cm}$ from its center; (b) on the axis of a disk of radius $10.0\ \text{cm}$ with $\sigma = 5.0\ \mu\text{C/m}^2$, $2.0\ \text{cm}$ from its center; and (c) at the center and at the surface of an insulating ball of radius $5.0\ \text{cm}$ carrying $+20\ \text{nC}$ uniformly.

Solution

(a) From equation (18.11), with $\sqrt{x^2 + a^2} = 0.050\ \text{m}$,

$$V = \frac{(8.99\times10^9)(12\times10^{-9})}{0.050} = \boxed{2.2\times10^3\ \text{V}}. \tag{18.4.7}$$

(b) From equation (18.13), with $\sigma/2\epsilon_0 = 2.82\times10^5\ \text{V/m}$,

$$V = (2.82\times10^5)\left(\sqrt{(0.020)^2 + (0.100)^2} - 0.020\right) = \boxed{2.3\times10^4\ \text{V}}. \tag{18.4.8}$$

(c) At the surface, $V = k_eQ/R = (8.99\times10^9)(20\times10^{-9})/0.050 = \boxed{3.6\times10^3\ \text{V}}$, and at the center it is 1.5 times larger, $\boxed{5.4\times10^3\ \text{V}}$. These are the same ring, disk, and ball whose fields were found in Chapter 16.

Checkpoint 18.5

A hollow metal sphere and a solid plastic sphere of the same radius carry the same total charge, spread uniformly over the metal sphere's surface and through the plastic sphere's volume. (a) Compare their potentials at a point outside. (b) Compare their potentials at the center. (c) Which has the stronger field at a point halfway from the center to the surface?

Answer

Answers to Checkpoints

Checkpoint 18.5. (a) Outside, both behave like a point charge Q at the center, so their potentials are equal: k_eQ/r. (b) The plastic sphere's center is at the higher potential. The metal sphere's interior is at the surface potential, k_eQ/R, since its interior field is zero. For the plastic sphere the field inside is not zero, so the potential keeps rising as you move inward, reaching $3k_eQ/2R$ at the center. (c) The plastic sphere. Inside the metal sphere the field is zero; inside the plastic sphere, at $r = R/2$, it is $k_eQ(R/2)/R^3 = k_eQ/(2R^2)$.

This page titled 18.4: Potential of a Continuous Charge Arrangement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.5: Calculating the Field from the Potential

Equation (18.1) finds the potential from the field by integrating. The process can be run in reverse. For a small step $d\mathbf{s}$, the change in potential is $dV = -\mathbf{E}\cdot d\mathbf{s} = -E_s\ ds$, where E_s is the component of the field along the step. Dividing by ds,

$$E_s = -\frac{dV}{ds}. \tag{18.5.1}$$

The component of the field in any direction is the negative of the rate at which the potential changes in that direction. Taking the steps along the three coordinate axes gives the components of the field:

Key Equation: Field from potential

$$E_x = -\frac{\partial V}{\partial x}, \qquad E_y = -\frac{\partial V}{\partial y}, \qquad E_z = -\frac{\partial V}{\partial z}. \tag{18.16}$$

The partial derivative $\partial V/\partial x$ means the ordinary derivative of V with respect to x, treating y and z as constants. In vector notation the three equations are written $\mathbf{E} = -\nabla V$, where ∇V, the **gradient** of V, is the vector that points in the direction in which V increases fastest, with a magnitude equal to that rate of increase. The field therefore points in the direction in which the potential *decreases* fastest, perpendicular to the equipotentials, as Section 18.2 found.

This gives a two-step strategy that is often the easiest way to find a field: first find the scalar potential by an integral that needs no components, then differentiate it. Every field computed in Chapters 15 and 16 can be checked this way.

- *Point charge:* $E_r = -\mathrm{d}(k_e q/r)/\mathrm{d}r = k_e q/r^2$.
- *Ring:* differentiating $V = k_e Q(x^2+a^2)^{-1/2}$ gives $E_x = k_e Q x/(x^2+a^2)^{3/2}$, which is equation (16.11).
- *Disk:* differentiating equation (18.13) gives $E_x = (\sigma/2\epsilon_0)\left(1 - x/\sqrt{x^2+R^2}\right)$, which is equation (16.12).
- *Dipole, on its axis:* with $\theta = 0$, $V = k_e p/r^2$ and $E_r = -\mathrm{d}V/\mathrm{d}r = 2k_e p/r^3$, which is equation (16.7).

Example 18.11: The field from a potential in two dimensions

In a certain region, the potential is $V(x,y) = C(3x^2y - y^3)$, with $C = 200\ \mathrm{V/m^3}$ and x and y in meters. Find the electric field at the point $(1.0\ \mathrm{m},\ 2.0\ \mathrm{m})$.

Solution

Differentiate, holding the other variable constant:

$$E_x = -\frac{\partial V}{\partial x} = -6Cxy, \qquad E_y = -\frac{\partial V}{\partial y} = -C(3x^2 - 3y^2). \tag{18.5.2}$$

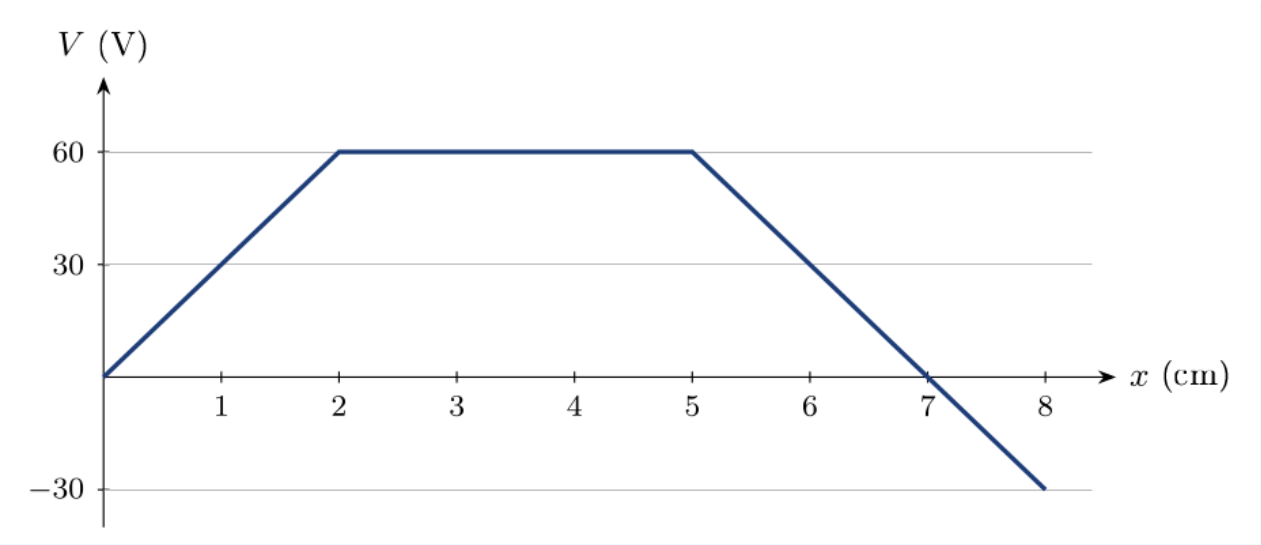
At $(1.0,\ 2.0)$,

$$E_x = -6(200)(1.0)(2.0) = -2400\,\mathrm{V/m}, \qquad E_y = -200(3-12) = +1800\,\mathrm{V/m}, \tag{18.5.3}$$

so $\mathbf{E} = \boxed{(-2.4\,\hat{\imath} + 1.8\,\hat{\jmath})\times 10^3\,\mathrm{V/m}}$, with magnitude $\boxed{3.0\times 10^3\,\mathrm{V/m}}$.

Example 18.12: Reading the field from a graph of potential

The graph shows the potential along the x-axis in a certain region. Find E_x in each of the three regions.



•Figure description
A graph of potential V in volts against position x in centimeters from 0 to 8. The potential rises in a straight line from 0 volts at x equals 0 to 60 volts at x equals 2 centimeters, stays flat at 60 volts from 2 to 5 centimeters, and then falls in a straight line to minus 30 volts at x equals 8 centimeters.

Solution

In each region the graph is a straight line, so $E_x = -\Delta V/\Delta x$ is constant.

- From 0 to 2 cm: $E_x = -(60-0)/(0.020) = \boxed{-3.0\times10^3\ \mathrm{V/m}}$. The field points in the $-x$ direction, back toward lower potential.
- From 2 to 5 cm: the potential is constant, so $\boxed{E_x = 0}$.
- From 5 to 8 cm: $E_x = -(-30-60)/(0.030) = \boxed{+3.0\times10^3\ \mathrm{V/m}}$.

A steep slope means a strong field, and the field always points “downhill.” The flat region could be the inside of a conductor.

Checkpoint 18.6

(a) At a certain point the potential is zero. Must the field be zero there? Give an example. (b) In a certain region the field is zero. What can you say about the potential there?

Answer

Answers to Checkpoints

Checkpoint 18.6. (a) No. The field depends on how the potential *changes* from point to point, not on its value. The midpoint of an electric dipole is an example: the potentials of the two charges cancel there, so $V = 0$, but their fields both point from $+$ toward $-$ and add, so $E \neq 0$. (b) The potential is constant throughout that region, but not necessarily zero. The interior of a charged conductor is an example: $E = 0$ inside, and the potential has the same nonzero value everywhere in it.

This page titled 18.5: Calculating the Field from the Potential is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.S: Electric Potential (Summary)

Summary

- The electric force is conservative, so it has a potential energy. The **potential difference** between two points is the change in potential energy per unit charge, $\Delta V = \Delta U/q_0 = -\int \mathbf{E}\cdot d\mathbf{s}$, measured in volts ($1\ \text{V} = 1\ \text{J/C}$; $1\ \text{V/m} = 1\ \text{N/C}$). In a uniform field $\Delta V = -Ed$. The field points toward lower potential.
- A charge q at potential V has **electric potential energy** $U = qV$. With only electric forces acting, $\Delta K = -q\ \Delta V$: positive charges accelerate toward lower potential, negative charges toward higher. One **electron volt** is 1.602×10^{-19} J.
- **Equipotential surfaces** are perpendicular to the field and crowd together where it is strong. A conductor in equilibrium is an equipotential; its surface charge density, and the field just outside, are greatest where it curves most sharply ($\sigma \propto 1/R$). An empty cavity in a conductor carries no charge and has no field.
- A point charge has $V = k_e q/r$ (zero at infinity), and a dipole far away has $V = k_e p\cos\theta/r^2$. Potentials of several charges add as scalars, $V = \sum k_e q_i/r_i$. Two charges have potential energy $U = k_e q_1 q_2/r$; a system's energy is the sum over pairs.
- For continuous distributions $V = k_e \int dq/r$: a ring gives $k_e Q/\sqrt{x^2+a^2}$ on its axis and a disk $(\sigma/2\epsilon_0)(\sqrt{x^2+R^2} - x)$. The potential can also be found by integrating a known field, as for the sphere and the infinite line.
- The field follows from the potential: $E_x = -\partial V/\partial x$ and so on, or $\mathbf{E} = -\nabla V$.

Key Equations

Potential difference

$$\Delta V = \frac{\Delta U}{q_0} = -\int_A^B \mathbf{E}\cdot d\mathbf{s}, \qquad \Delta V = -Ed \quad \text{(uniform field)} \tag{18.S.1}$$

Potential energy of a charge, and energy conservation

$$U = qV, \qquad \Delta K = -q\,\Delta V, \qquad 1\,\text{eV} = 1.602 \times 10^{-19}\,\text{J} \tag{18.S.2}$$

Point charge, and system of point charges

$$V = \frac{k_e q}{r}, \qquad V = \sum_i \frac{k_e q_i}{r_i} \tag{18.S.3}$$

Electric dipole (far away)

$$V = \frac{k_e p\cos\theta}{r^2} \tag{18.S.4}$$

Potential energy of two point charges

$$U = \frac{k_e q_1 q_2}{r} \tag{18.S.5}$$

Continuous distribution

$$V = \int dV = k_e \int \frac{dq}{r} \tag{18.S.6}$$

Ring and disk (on the axis)

$$V_{\text{ring}} = \frac{k_e Q}{\sqrt{x^2 + a^2}}, \qquad V_{\text{disk}} = \frac{\sigma}{2\epsilon_0}\left(\sqrt{x^2 + R^2} - x\right) \tag{18.S.7}$$

Field from potential

$$E_x = -\frac{\partial V}{\partial x}, \quad E_y = -\frac{\partial V}{\partial y}, \quad E_z = -\frac{\partial V}{\partial z}; \qquad \mathbf{E} = -\nabla V \tag{18.S.8}$$

This page titled 18.S: Electric Potential (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

18.E: Electric Potential (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $k_e = 8.99 \times 10^9\ \text{N}\cdot\text{m}^2/\text{C}^2$, $\epsilon_0 = 8.85 \times 10^{-12}\ \text{C}^2/(\text{N}\cdot\text{m}^2)$, $e = 1.602 \times 10^{-19}\ \text{C}$, $m_e = 9.11 \times 10^{-31}\ \text{kg}$, and $m_p = 1.673 \times 10^{-27}\ \text{kg}$.

Section 18.1 — Electric Potential

18.1 ★ A uniform field of $800\ \text{V/m}$ points in the $+y$ direction. Find $V_B - V_A$ when B is (a) $0.25\ \text{m}$ above A along y, (b) $0.25\ \text{m}$ to the right of A along x, and (c) $0.25\ \text{m}$ below A.

18.2 ★ *[Real-world: atmospheric physics.]* The fair-weather electric field near the ground is about $150\ \text{V/m}$, pointing down. Estimate the potential difference between the top of your head and your feet, taking your height to be about $1.8\ \text{m}$. Why don't you get a shock?

18.3 ★★ Two parallel plates $4.0\ \text{mm}$ apart are connected to a $12\ \text{V}$ battery. (a) Find the field between them. (b) How much work does the field do on an electron that moves from the negative plate to the positive plate?

18.4 ★★ *[Conceptual.]* A student claims that because the potential at a point is $0\ \text{V}$, no work is needed to bring a charge there from far away, and therefore there can be no force on the charge when it arrives. Which part of the claim is right, and which is wrong?

Section 18.1.1 — Electric Potential Energy

18.5 ★ How much kinetic energy, in electron volts and in joules, does an alpha particle (charge $+2e$) gain when it is accelerated from rest through a potential difference of $2.0\ \text{kV}$?

18.6 ★★ *[Real-world: medical imaging.]* In an X-ray tube, electrons are accelerated from rest through $50\ \text{kV}$ before striking a metal target. Find (a) their kinetic energy in keV and in joules and (b) their final speed, ignoring relativity. (c) Compare the speed with the speed of light, and comment on whether ignoring relativity was reasonable.

18.7 ★★ An electron moving at $5.0 \times 10^6\ \text{m/s}$ enters a region where it moves toward lower potential. Through what potential difference can it travel before it stops?

18.8 ★★ A proton is released from rest at point A, where the potential is $+450\ \text{V}$. It moves to point B, where the potential is $+150\ \text{V}$. Find its speed at B. What would happen to an electron released from rest at A?

Section 18.2 — Equipotential Surfaces

18.9 ★ *[Conceptual.]* Sketch the equipotentials and field lines of two equal positive charges (compare Figure 16.4). Where is the potential zero, if anywhere?

18.10 ★★ Equipotential surfaces in a region are parallel planes, $5.0\ \text{mm}$ apart, at $10\ \text{V}$, $20\ \text{V}$, $30\ \text{V}$, and so on, with the potential increasing toward the east. Find the magnitude and direction of the field.

18.11 ★★ A conducting sphere of radius $15\ \text{cm}$ is charged to a potential of $12\ \text{kV}$. Find its charge, and the field just outside its surface. Air breaks down at $3.0 \times 10^6\ \text{V/m}$; what is the highest potential this sphere could reach?

18.12 ★★★ Two conducting spheres, of radii $6.0\ \text{cm}$ and $3.0\ \text{cm}$, are far apart. The larger carries $+12\ \text{nC}$ and the smaller is uncharged. They are then connected by a long thin wire. (a) Find the charge on each sphere afterward. (b) Find the potential of each sphere before and after. (c) Find the ratio of their surface charge densities afterward.

Section 18.3 — Potential of a Discrete Charge Arrangement

18.13 ★ Find the potential $20\ \text{cm}$ from a point charge of $-4.0\ \text{nC}$. How much work is needed to bring a $+1.0\ \text{nC}$ charge from very far away to that point?

18.14 ★ At what distance from a $+1.0\ \mu\text{C}$ point charge is the potential $1000\ \text{V}$? Does the potential change by the same amount between $1\ \text{m}$ and $2\ \text{m}$ as between $2\ \text{m}$ and $3\ \text{m}$?

18.15 ★★ Charges of $+3.0\ \text{nC}$ and $-6.0\ \text{nC}$ sit $0.30\ \text{m}$ apart on the x-axis, the positive one at the origin. Find all points on the x-axis where the potential is zero.

18.16 ★★ Four charges sit at the corners of a square of side $0.20\ \text{m}$: $+q$, $+q$, $-q$, and $-q$ going around the square, with $q = 5.0\ \text{nC}$. (a) Find the potential at the center. (b) Find the potential energy of the system.

18.17 ★★ An electric dipole has charges $\pm 4.0\ \text{nC}$ separated by $2.0\ \text{mm}$. Find the potential $10\ \text{cm}$ from its center (a) on its axis, on the positive side, (b) at 45° from its axis, and (c) on its perpendicular bisector.

18.18 ★★ *[Real-world: nuclear physics.]* In a fission event, a uranium nucleus splits into two fragments with charges $+46e$ and $+46e$ whose centers are initially $1.3 \times 10^{-14}\ \text{m}$ apart. (a) Find their electric potential energy, in MeV. (b) If all of it becomes kinetic energy as they fly apart, what is each fragment's kinetic energy? (Take the two fragments to have equal masses.)

18.19 ★★★ Three charges, $+q$, $+q$, and $-q$, are to be placed at the corners of an equilateral triangle of side a. (a) Find the potential energy of the arrangement. (b) Is positive or negative work needed to assemble it? (c) What would the answer be if all three charges were $+q$?

Section 18.4 — Potential of a Continuous Charge Arrangement

18.20 ★ A thin ring of radius $5.0\ \text{cm}$ carries $-8.0\ \text{nC}$. Find the potential at its center and $12\ \text{cm}$ from its center on its axis.

18.21 ★★ A thin rod of length L carries charge Q uniformly. Show that the potential at a point on its axis, a distance d beyond one end, is $V = (k_e Q/L)\ln\big((d+L)/d\big)$. Check that it approaches $k_e Q/d$ when $d \gg L$.

18.22 ★★ *[Real-world: electronics.]* A long coaxial cable has an inner wire of radius $0.50\ \text{mm}$ carrying $+5.0\ \text{nC/m}$ and an outer conducting sheath of radius $3.0\ \text{mm}$ carrying $-5.0\ \text{nC/m}$. Find the potential difference between the wire and the sheath. Which is at the higher potential?

18.23 ★★ An insulating sphere of radius $8.0\ \text{cm}$ carries charge $30\ \text{nC}$ spread uniformly through its volume. Find the potential at its center, at $r = 4.0\ \text{cm}$, at its surface, and at $r = 16\ \text{cm}$.

18.24 ★★★ A thin disk of radius R carries uniform surface charge density σ. (a) Find the potential at its center. (b) Use a binomial expansion to show that far away on the axis, $x \gg R$, equation (18.13) becomes $k_e Q/x$ with $Q = \sigma\pi R^2$.

Section 18.5 — Calculating the Field from the Potential

18.25 ★ In a region, the potential along the x-axis is $V(x) = (250\ \text{V/m}^2)x^2 - (40\ \text{V/m})x$. Find E_x at $x = 0$, at $x = 0.080\ \text{m}$, and at $x = 0.20\ \text{m}$. Where is the field zero?

18.26 ★★ The potential in a region is $V(x,y,z) = (3.0\ \text{V/m}^2)\ xy - (5.0\ \text{V/m}^2)\ z^2$. Find the electric field at the point $(2.0\ \text{m},\ 1.0\ \text{m},\ -1.0\ \text{m})$, and its magnitude.

18.27 ★★ *[Graph-reading.]* Use the graph in Example 18.12. (a) Sketch E_x against x. (b) An electron is released at rest at $x = 3.5\ \text{cm}$. Which way does it move once nudged slightly to the right? (c) What kinetic energy does it have, in eV, when it reaches $x = 8.0\ \text{cm}$ if it does?

18.28 ★★★ Starting from the dipole potential written in Cartesian coordinates, $V = k_e p z/(x^2+y^2+z^2)^{3/2}$, with the dipole along z, find E_z on the perpendicular bisector ($z = 0$) and show that it agrees with equation (16.7).

18.29 ★★★ *[Real-world: atomic physics.]* In the Bohr model of hydrogen the electron circles the proton at radius r with the Coulomb force supplying the centripetal force. (a) Show that the electron's kinetic energy is $K = k_e e^2/2r$ and its potential energy is $U = -k_e e^2/r$, so that $K = -U/2$. (b) Find the total energy at $r = 5.29 \times 10^{-11}\ \text{m}$, in eV, and explain why this is the ionization energy of hydrogen.

18.30 ★★★ A uniformly charged insulating sphere of radius R and charge $+Q$ has a narrow tunnel through its center. An electron is released from rest at the surface. Using equation (18.14), find its speed as it passes through the center. Check your answer against Problem 16.31 (the same motion described with forces).

18.31 ★★★ *[Real-world: laboratory equipment.]* A Van de Graaff generator's sphere of radius R is surrounded by air that breaks down at $E_{\text{max}} = 3.0 \times 10^6\ \text{V/m}$. (a) Show that the highest potential the sphere can reach is $V_{\text{max}} = E_{\text{max}} R$. (b) How large must the sphere be to reach $1.0\ \text{MV}$? (c) Explain, using Section 18.2, why the sphere must be smooth, with no sharp edges or protruding screws.

This page titled 18.E: Electric Potential (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19: CAPACITANCE

A camera flash charges for a second or two, then releases its stored energy in a thousandth of a second, far faster than any battery could deliver it. A heart defibrillator does the same on a larger scale, storing a couple of hundred joules and then driving them through a patient's chest in a few milliseconds. The touchscreen on your phone senses your finger because your finger changes how much charge a grid of tiny electrodes can hold. And every bit in your computer's working memory is roughly a hundred thousand electrons, stored or not stored, on a microscopic conductor.

All of these devices are **capacitors**: two conductors, close together but not touching, that store separated charge and, with it, electrical energy. This chapter asks how much charge a pair of conductors can hold for a given voltage, where the stored energy actually resides, how filling the gap with an insulator changes things, and how capacitors combine when they are connected together. The answers draw on everything in the last three chapters: the fields of Chapter 16, the Gauss's-law results of Chapter 17, and the potential of Chapter 18.

Learning Objectives

- Define capacitance, and calculate it for parallel-plate, cylindrical, and spherical capacitors.
- Find the energy stored in a capacitor, and the energy density of an electric field.
- Explain how a dielectric increases capacitance, and apply the dielectric constant and dielectric strength.
- Find the equivalent capacitance of capacitors in parallel and in series, and the charge on and voltage across each.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

19.1: Capacitance
19.2: Energy Stored in an Electric Field
19.3: Dielectrics
19.4: Capacitors in Parallel and in Series
19.S: Capacitance (Summary)
19.E: Capacitance (Problems)

This page titled 19: Capacitance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.1: Capacitance

A **capacitor** is any pair of conductors separated by an insulator or by empty space. To charge it, electrons are moved from one conductor to the other, typically by connecting the two to the terminals of a battery. One conductor ends up with charge $+Q$ and the other with $-Q$; we call Q the charge *on the capacitor*, even though its net charge is zero. The separated charges produce a field between the conductors and so a potential difference ΔV between them. Because every field in electrostatics is proportional to the charges producing it, ΔV is proportional to Q. Double the charge and the field doubles everywhere, and so does the potential difference. The constant of proportionality is the **capacitance**:

Key Equation: Capacitance

$$C = \frac{Q}{\Delta V}. \tag{19.1}$$

Capacitance measures how much charge the capacitor holds per volt. It is always positive, and it depends only on the shapes, sizes, and arrangement of the conductors and on the material between them, not on Q or ΔV. For a capacitor, we will usually write the potential difference simply as V, so that $Q = CV$.

The SI unit of capacitance is the **farad** (F), after Michael Faraday: $1\ \mathrm{F} = 1\ \mathrm{C/V}$. As we will see, a farad is an enormous capacitance. Capacitors in electronic circuits typically range from picofarads ($1\ \mathrm{pF} = 10^{-12}\ \mathrm{F}$) to microfarads ($1\ \mu\mathrm{F} = 10^{-6}\ \mathrm{F}$).

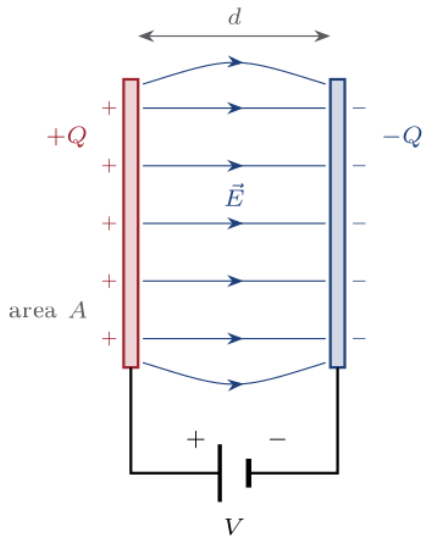


Figure 19.1. A parallel-plate capacitor charged by a battery. The battery moves charge until the potential difference between the plates equals its own voltage V, leaving $\pm Q$ on the plates with $Q = CV$. Between large plates the field is uniform; only near the edges does it fringe outward.

•Figure description

A parallel-plate capacitor connected to a battery. Two vertical plates of area A are a distance d apart. The left plate, connected to the battery's positive terminal, carries positive charge; the right plate, connected to the negative terminal, carries negative charge. Evenly spaced horizontal field lines run from the positive to the negative plate between them, with slight fringing at the edges. Wires run from each plate down to a battery symbol below labeled V.

The parallel-plate capacitor. The simplest and most important capacitor is a pair of parallel conducting plates, each of area A, a distance d apart (Figure 19.1). With charges $\pm Q$, each plate has surface charge density $\sigma = Q/A$ on its inner face. If the plates are large compared with their separation, the field between them is uniform and equal to σ/ϵ_0 (equation (16.14)), so

$$E = \frac{Q}{\epsilon_0 A}, \qquad V = Ed = \frac{Qd}{\epsilon_0 A}, \tag{19.1.1}$$

and the capacitance is

Key Equation: Parallel-plate capacitor

$$C = \frac{\epsilon_0 A}{d}. \tag{19.2}$$

Larger plates hold more charge at a given voltage, and closer plates do too, because a given voltage then produces a stronger field. This result also gives ϵ_0 a convenient unit: since $C = \epsilon_0 A/d$, $\epsilon_0 = 8.85 \times 10^{-12}\ \mathrm{F/m}$, the same thing as $\mathrm{C^2/(N \cdot m^2)}$.

Charging a capacitor with a battery. A battery maintains a fixed potential difference between its terminals (Chapter 21). Connected to an uncharged capacitor, it pulls electrons off the plate attached to its positive terminal and pushes them onto the plate attached to its negative terminal. As the charge builds up, so does the capacitor's own potential difference, and the flow stops when the two are equal. The capacitor then holds charge $Q = CV$, where V is the battery's voltage. If the battery is disconnected, the charge has nowhere to go and stays on the plates.

Example 19.1: How big is a farad?

(a) Find the capacitance of two square plates, $10.0\ \mathrm{cm}$ on a side, $1.00\ \mathrm{mm}$ apart in air. (b) How much charge does this capacitor hold when connected to a $12.0\ \mathrm{V}$ battery? (c) How large would the plates of a $1.00\ \mathrm{F}$ capacitor with the same spacing have to be?

Solution

(a) From equation (19.2),

$$C = \frac{\epsilon_0 A}{d} = \frac{(8.85 \times 10^{-12})(0.100)^2}{1.00 \times 10^{-3}} = \boxed{8.85 \times 10^{-11}\,\mathrm{F}} = 88.5\,\mathrm{pF}. \tag{19.1.2}$$

(b) $Q = CV = (8.85 \times 10^{-11})(12.0) = \boxed{1.06 \times 10^{-9}\,\mathrm{C}}$, about a nanocoulomb.

(c) Solving for A with $C = 1.00\ \mathrm{F}$,

$$A = \frac{Cd}{\epsilon_0} = \frac{(1.00)(1.00 \times 10^{-3})}{8.85 \times 10^{-12}} = \boxed{1.13 \times 10^{8}\,\mathrm{m^2}}, \tag{19.1.3}$$

a square more than ten kilometers on a side. A farad is a huge capacitance. The "supercapacitors" now sold with capacitances of many farads get there by using a separation measured in nanometers and an effective plate area of thousands of square meters per gram of porous carbon.

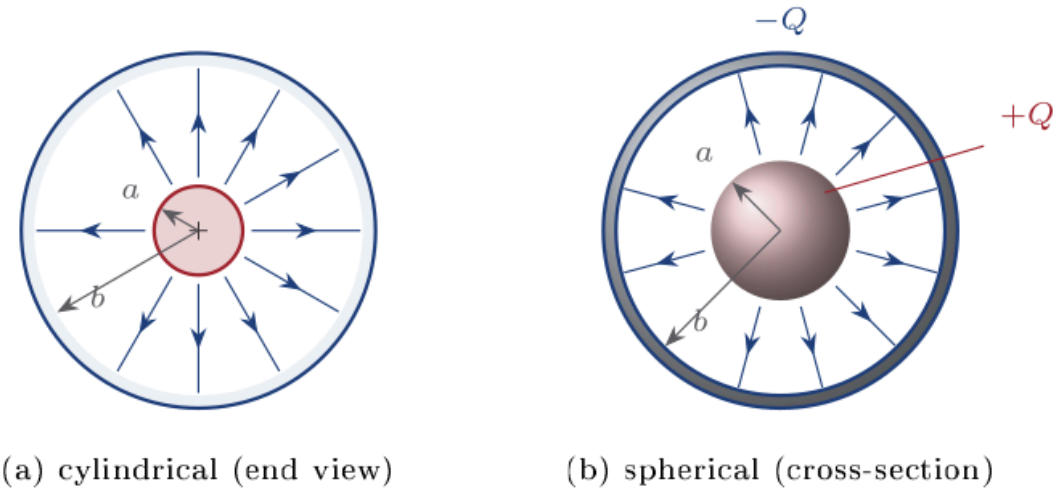


Figure 19.2. (a) A cylindrical capacitor, such as a coaxial cable, seen end-on: an inner conductor of radius a and an outer conductor of radius b. (b) A spherical capacitor: concentric spheres of radii a and b. In both, the field between the conductors points radially from the positive conductor to the negative one.

•Figure description

Two panels. Panel a shows a cross-section of a cylindrical capacitor: a small central circle of radius a, the inner conductor carrying positive charge, surrounded by a larger concentric circle of radius b, the outer conductor carrying negative charge, with radial field lines between them pointing outward. Panel b shows a cross-section of a spherical capacitor: an inner sphere of radius a with positive charge inside an outer spherical shell of radius b with negative charge, again with radial field lines between them.

The cylindrical capacitor. A coaxial cable is a capacitor: a central wire of radius a surrounded by a conducting cylinder of radius b (Figure 19.2a). Give the inner conductor charge $+Q$ and the outer $-Q$, spread over a length L so that $\lambda = Q/L$. Between them the field is that of a line of charge, and Chapter 18 found the potential difference, equation (18.15): $V = (\lambda/2\pi\epsilon_0)\ln(b/a)$, with the inner conductor higher. Then $C = Q/V = \lambda L/V$, or

Key Equation: Cylindrical capacitor

$$C = \frac{2\pi\epsilon_0 L}{\ln(b/a)}. \tag{19.3}$$

Once again the capacitance depends only on geometry. The capacitance per unit length, C/L, is an important specification of any coaxial cable.

The spherical capacitor. Two concentric conducting spheres of radii a and $b > a$ carry charges $+Q$ and $-Q$ (Figure 19.2b). Between them the field is that of a point charge, k_eQ/r^2, so the potential difference is

$$V = \int_a^b \frac{k_eQ}{r^2}\,dr = k_eQ\left(\frac{1}{a} - \frac{1}{b}\right) = k_eQ\,\frac{b-a}{ab}, \tag{19.1.4}$$

and

Key Equation: Spherical capacitor

$$C = \frac{ab}{k_e(b-a)} = 4\pi\epsilon_0\frac{ab}{b-a}. \tag{19.4}$$

Let the outer sphere grow without limit, $b \to \infty$. Then $b/(b-a) \to 1$, and the capacitance approaches $C = 4\pi\epsilon_0 a$. This is the capacitance of an *isolated* sphere, whose "other plate" is effectively at infinity: charge Q on it raises it to potential $V = k_eQ/a$, so $C = Q/V = a/k_e$.

Example 19.2: The capacitance of a coaxial cable and of the Earth

(a) A coaxial cable has an inner wire of radius $0.50\ \text{mm}$ and an outer conductor of radius $3.0\ \text{mm}$, with air between them. Find its capacitance per meter. (b) Treating the Earth as an isolated conducting sphere of radius $6.37 \times 10^6\ \text{m}$, find its capacitance.

Solution

(a) From equation (19.3),

$$\frac{C}{L} = \frac{2\pi\epsilon_0}{\ln(b/a)} = \frac{2\pi(8.85 \times 10^{-12})}{\ln 6.0} = \boxed{3.1 \times 10^{-11}\,\text{F/m}} = 31\,\text{pF/m}. \tag{19.1.5}$$

Real coaxial cables, filled with plastic rather than air, have two or three times more, around $70\text{–}100\ \text{pF/m}$, for a reason Section 19.3 explains.

(b) For an isolated sphere,

$$C = 4\pi\epsilon_0 R = 4\pi(8.85 \times 10^{-12})(6.37 \times 10^6) = \boxed{7.1 \times 10^{-4}\,\text{F}} = 710\,\mu\text{F}. \tag{19.1.6}$$

The whole planet has less capacitance than a single electrolytic capacitor from an electronics catalog. Capacitance grows only in proportion to size, while bringing two conductors close together, as a capacitor does, pays off far more.

Checkpoint 19.1

A parallel-plate capacitor is connected to a battery. Without disconnecting the battery, you pull the plates apart to twice their original separation. (a) What happens to the capacitance? (b) To the potential difference? (c) To the charge on the plates? (d) Where does the charge go?

Answer

Answers to Checkpoints

Checkpoint 19.1. (a) $C = \epsilon_0 A/d$, so doubling d halves the capacitance. (b) The potential difference stays the same: the battery holds it fixed as long as it is connected. (c) With $Q = CV$, halving C at fixed V halves the charge. (d) The charge that leaves flows back through the wires into the battery, against the battery's emf.

This page titled 19.1: Capacitance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.2: Energy Stored in an Electric Field

Charging a capacitor takes work, because each bit of charge moved from one plate to the other must be pushed against the field that the charge already there has built up. That work is stored as electric potential energy, and it can be recovered, which is what the camera flash and the defibrillator do.

Suppose that partway through the charging, the plates carry charge q, so the potential difference between them is $v = q/C$. Moving a further small charge dq from the negative plate to the positive plate raises its potential energy by $dU = v\ dq = (q/C)\ dq$. Adding up the work from $q = 0$ to the final charge Q,

$$U = \int_0^Q \frac{q}{C}\, dq = \frac{Q^2}{2C}. \tag{19.2.1}$$

Using $Q = CV$, the **energy stored in a capacitor** can be written in three equivalent ways:

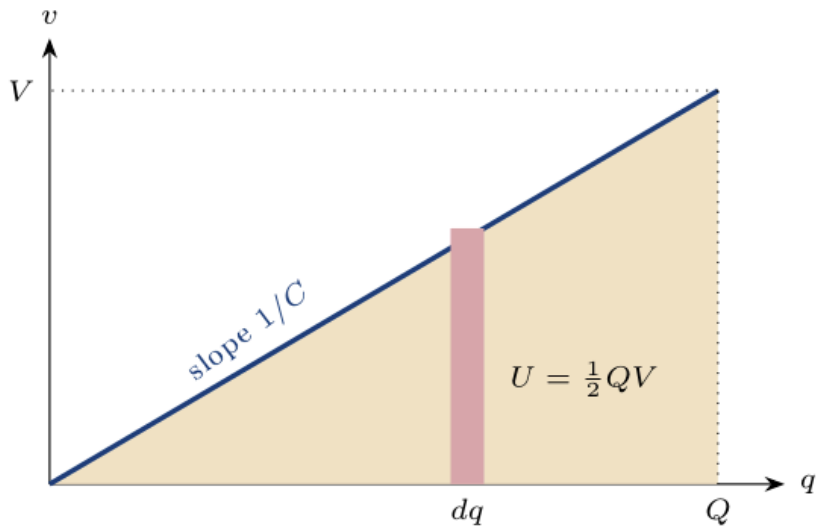


Figure 19.3. The energy stored in a capacitor is the area under the graph of potential difference against charge. Moving charge dq at potential difference v takes work $v\ dq$ (the red strip); the total is the triangle, $\frac{1}{2}QV$.

•Figure description

A graph of the capacitor's potential difference v, on the vertical axis, against the charge q on its plates, on the horizontal axis. The graph is a straight line through the origin with slope one over C, reaching V at the final charge Q. The triangle under the line is shaded, and its area, one half Q V, is the stored energy. A thin vertical strip of width d q under the line at charge q represents the small amount of work v d q done to move the next bit of charge.

Key Equation: Energy stored in a capacitor

$$U = \frac{Q^2}{2C} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV. \tag{19.5}$$

Why $\frac{1}{2}QV$ and not QV? Because the charge was not all moved through the final potential difference V. The first bit of charge was moved through almost no potential difference at all, and the last bit through the full V. On average, the charge was moved through $V/2$. On a graph of v against q, the stored energy is the area of the triangle under the straight line (Figure 19.3), just as the work done stretching a spring was the area under a graph of force against stretch in Chapter 7.

The energy is in the field. Where is this energy? Apply equation (19.5) to a parallel-plate capacitor, with $C = \epsilon_0 A/d$ and $V = Ed$:

$$U = \tfrac{1}{2}\left(\frac{\epsilon_0 A}{d}\right)(Ed)^2 = \tfrac{1}{2}\epsilon_0 E^2\,(Ad). \tag{19.2.2}$$

The product Ad is the volume of the space between the plates, where the field is. The energy per unit volume, the **energy density** of the field, is therefore

Key Equation: Energy density of an electric field

$$u_E = \tfrac{1}{2}\epsilon_0 E^2. \tag{19.6}$$

Although we derived it for the uniform field of a parallel-plate capacitor, equation (19.6) holds for any electric field in vacuum. It invites a change of viewpoint: the energy of a charged capacitor is not stored "in the charges" but *in the field* between them, wherever that field is. The viewpoint becomes unavoidable in Chapter 27, where electromagnetic waves carry energy across empty space, far from any charge, in their electric and magnetic fields.

Example 19.3: A camera flash and a defibrillator

(a) A camera flash unit charges a $150\ \mu\text{F}$ capacitor to $300\ \text{V}$. How much energy does it store, and what average power does the flash deliver if the energy is released in $1.0\ \text{ms}$? (b) A defibrillator must deliver $200\ \text{J}$ from a capacitor charged to $5.0\ \text{kV}$. What capacitance does it need, and how much charge passes through the patient?

Solution

(a) From equation (19.5),

$$U = \tfrac{1}{2}CV^2 = \tfrac{1}{2}(150\times10^{-6})(300)^2 = \boxed{6.8\,\text{J}}, \tag{19.2.3}$$

and the average power is $P = U/t = 6.8/(1.0\times10^{-3}) = \boxed{6.8\,\text{kW}}$. The batteries charging the capacitor supply energy slowly, over a second or two; the capacitor delivers it a thousand times faster.

(b) Solving $U = \frac{1}{2}CV^2$ for C,

$$C = \frac{2U}{V^2} = \frac{2(200)}{(5.0\times10^3)^2} = \boxed{16\,\mu\text{F}}, \tag{19.2.4}$$

and the charge is $Q = CV = (16\times10^{-6})(5.0\times10^3) = \boxed{0.080\,\text{C}}$.

Example 19.4: Energy density near a thundercloud

Find the energy density of the electric field (a) near the ground on a fair day, $E = 150\ \text{V/m}$; and (b) in air just about to break down, $E = 3.0\times10^6\ \text{V/m}$. (c) Estimate the energy stored in the field between a thundercloud and the ground, if the field of case (b) fills a region $1\ \text{km}$ high and $2\ \text{km}$ across.

Solution

(a) $u_E = \frac{1}{2}\epsilon_0 E^2 = \frac{1}{2}(8.85\times10^{-12})(150)^2 = \boxed{1.0\times10^{-7}\,\text{J/m}^3}$.

(b) $u_E = \frac{1}{2}(8.85\times10^{-12})(3.0\times10^6)^2 = \boxed{40\,\text{J/m}^3}$.

(c) The volume is roughly $(1000\ \text{m})(2000\,\text{m})^2 = 4\times10^9\ \text{m}^3$, so $U \approx (40)(4\times10^9) \approx \boxed{2\times10^{11}\,\text{J}}$, the energy of some forty tons of TNT. In reality the field is strong only in parts of this region, and a single lightning stroke releases about 10^9 joules. But the estimate shows why a thunderstorm is one of the most energetic events in the atmosphere.

Example 19.5: Pulling the plates apart

A parallel-plate capacitor of capacitance C_0 is charged to potential difference V_0 and then disconnected from the battery. The plates are then pulled apart to twice their original separation. Find the new charge, potential difference, and stored energy, and account for the change in energy.

Solution

Disconnected, the capacitor keeps its charge: $Q = Q_0 = C_0V_0$. Doubling d halves the capacitance, $C = C_0/2$, so the potential difference doubles, $V = Q/C = 2V_0$. The field $E = Q/\epsilon_0 A$ is unchanged, since the charge on each plate is. The stored energy is

$$U = \frac{Q^2}{2C} = \frac{Q_0^2}{2(C_0/2)} = 2U_0. \tag{19.2.5}$$

The energy has *doubled*. The same field now fills twice the volume, so equation (19.6) gives twice the energy too. The extra energy came from the work you did pulling the plates apart against their mutual attraction: opposite charges attract, and separating them takes work.

Checkpoint 19.2

Repeat Example 19.5, but with the battery still connected while the plates are pulled apart. (a) What happens to V, C, Q, and U ? (b) The stored energy now *decreases*, even though you still do positive work pulling the plates apart. Where does the energy go? *(Hint: what happens to the charge that leaves the plates?)*

Answer

Answers to Checkpoints

Checkpoint 19.2. (a) V stays fixed (the battery sets it); C halves; $Q = CV$ halves; and $U = \frac{1}{2}CV^2$ halves. (b) The charge driven off the plates is pushed back into the battery *against* its emf, which recharges the battery. The energy the battery gains equals your mechanical work *plus* the energy released by the capacitor's field. Nothing is lost: energy is transferred from you and from the field into the battery.

This page titled 19.2: Energy Stored in an Electric Field is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.3: Dielectrics

Most real capacitors are not empty. The space between the plates is filled with an insulating material such as plastic, ceramic, glass, paper, or oil, called a **dielectric**. In 1837 Faraday found that filling the space between a capacitor's plates with an insulator *increases* its capacitance, by a factor that depends only on the material. That factor is the **dielectric constant** κ (the Greek letter kappa) of the material:

Key Equation: Capacitance with a dielectric

$$C = \kappa C_0, \tag{19.7}$$

where C_0 is the capacitance of the same capacitor with vacuum between its plates. For vacuum $\kappa = 1$ exactly; for every material $\kappa > 1$. Table 19.1 lists some values. For a parallel-plate capacitor filled with a dielectric, $C = \kappa\epsilon_0 A/d$, and it is often convenient to combine the two constants into the **permittivity** of the material, $\epsilon = \kappa\epsilon_0$, so that $C = \epsilon A/d$.

A dielectric does a second job, just as important in practice. Every insulator breaks down and conducts if the field in it is made strong enough; the air in a spark gap does so at about 3×10^6 V/m. The largest field a material can withstand without breaking down is its **dielectric strength**. Most dielectrics are far stronger than air, so filling a capacitor with one allows a much higher voltage across the same gap. The two effects together are why nearly every practical capacitor contains a dielectric.

Table 19.1. Approximate dielectric constants and dielectric strengths of some materials at room temperature. Water's large κ comes from its permanently polar molecules; it is too conductive, in practice, to serve as a capacitor dielectric.

Material	Dielectric constant κ	Dielectric strength (V/m)
Vacuum	1 (exactly)	—
Air (dry, 1 atm)	1.00059	3×10^6
Paper	3.7	16×10^6
Teflon	2.1	60×10^6
Polystyrene	2.6	24×10^6
Pyrex glass	4.7	14×10^6
Mica	5.4	100×10^6
Water (pure, 20 °C)	80	—
Strontium titanate	~ 300	8×10^6

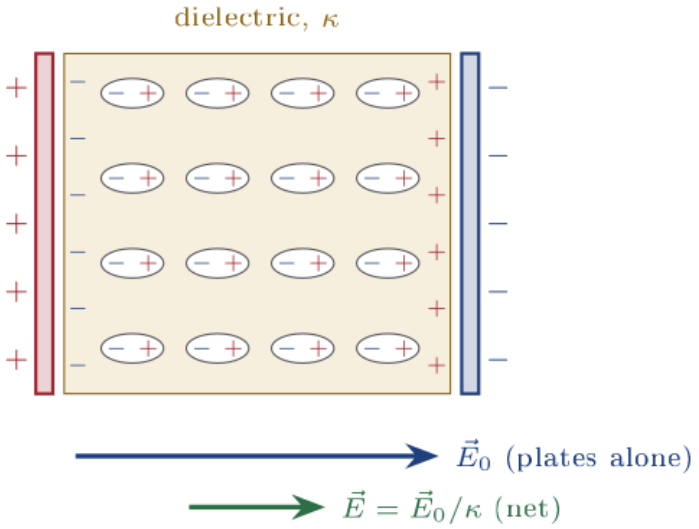


Figure 19.4. A dielectric in a capacitor. The field polarizes the molecules, leaving a layer of induced charge on each face of the slab, opposite in sign to the charge on the nearby plate. The induced charges partly cancel the field of the plates, reducing it to E_0/κ inside the dielectric.

•Figure description

A parallel-plate capacitor with a dielectric slab filling the space between its plates. The left plate is positive and the right plate negative. Inside the slab, small oval molecules are all polarized with their negative ends toward the positive plate and positive ends toward the negative plate. At the slab's left face a layer of negative induced charge faces the positive plate, and at its right face a layer of positive induced charge faces the negative plate. Below, two arrows compare the fields: the long arrow E sub zero from the plate charges, and a shorter arrow E equals E sub zero over kappa for the net field inside the dielectric.

How a dielectric works. The explanation lies in polarization, which we first met in Section 15.1. Place a slab of dielectric in the field between the plates (Figure 19.4). The field pulls the electron cloud of each molecule slightly one way and its nuclei the other, and if the molecules are already polar, like water, it tends to swing them into line (Section 16.5). Either way, the material fills up with tiny dipoles aligned with the field. Inside the material, the positive end of each dipole sits next to the negative end of its neighbor, and the charges cancel. But at the two surfaces they do not: a thin layer of negative *induced charge* appears on the face next to the positive plate, and a layer of positive induced charge on the face next to the negative plate. These induced charges produce a field of their own, pointing opposite to the field of the charges on the plates, and so the net field inside the dielectric is weaker:

$$E = \frac{E_0}{\kappa}. \tag{19.8}$$

With the same charge Q on the plates, a weaker field means a smaller potential difference, $V = Ed = V_0/\kappa$, and therefore a larger capacitance, $C = Q/V = \kappa C_0$.

Note: Gauss's law in a dielectric

The induced charges are real charges, and Gauss's law, $\oint \mathbf{E} \cdot d\mathbf{A} = q_{\text{enc}}/\epsilon_0$, holds with q_{enc} counting them. But they are tied to the molecules and cannot be controlled directly. For a uniform, linear dielectric, their effect is to reduce the field by exactly the factor κ everywhere, so Gauss's law can be rewritten using only the *free* charge, the charge we actually put on the conductors:

$$\oint \kappa\mathbf{E} \cdot d\mathbf{A} = \frac{q_{\text{free}}}{\epsilon_0}. \tag{19.3.1}$$

Every result of Chapters 15 through 18 therefore carries over to charges embedded in a dielectric by replacing ϵ_0 with $\epsilon = \kappa\epsilon_0$. Two charges in water, for instance, attract each other 80 times more weakly than in vacuum, which is why salt dissolves so readily in water: the ions' strong attraction is weakened enough for thermal motion to pull them apart.

Example 19.6: Inserting a dielectric, with and without the battery

A $50\ \text{pF}$ air-filled capacitor is charged by a $100\ \text{V}$ battery. A slab of Pyrex ($\kappa = 4.7$) is then slid in to fill the space between the plates. Find the new capacitance, charge, voltage, and stored energy (a) if the battery remains connected and (b) if the battery was disconnected first.

Solution

Before the slab: $Q_0 = C_0V_0 = (50 \times 10^{-12})(100) = 5.0\ \text{nC}$, and $U_0 = \frac{1}{2}C_0V_0^2 = 0.25\ \mu\text{J}$. In both cases the new capacitance is $C = \kappa C_0 = \boxed{235\,\text{pF}}$.

(a) With the battery connected, $V = \boxed{100\,\text{V}}$ is fixed. The battery supplies more charge, $Q = CV = \boxed{23.5\,\text{nC}}$, and the energy rises to $U = \frac{1}{2}CV^2 = \kappa U_0 = \boxed{1.2\,\mu\text{J}}$.

(b) With the battery disconnected, $Q = \boxed{5.0\,\text{nC}}$ is fixed. The voltage falls to $V = Q/C = V_0/\kappa = \boxed{21\,\text{V}}$, and the energy falls to $U = Q^2/2C = U_0/\kappa = \boxed{0.053\,\mu\text{J}}$.

In case (b) the capacitor loses energy as the slab goes in. The missing energy went into work done *on* the slab: the fringing field at the edge of the plates pulls a dielectric into the gap, just as a charged comb pulls on a scrap of paper. You would have to hold the slab back to keep it from being sucked in.

Example 19.7: The largest charge a capacitor can hold

A capacitor is made of two sheets of aluminum foil, each $0.40\ \text{m}^2$, separated by a sheet of polystyrene $0.050\ \text{mm}$ thick. Find (a) its capacitance, (b) the largest voltage it can withstand, and (c) the largest charge it can hold.

Solution

(a) With $\kappa = 2.6$,

$$C = \frac{\kappa\epsilon_0 A}{d} = \frac{(2.6)(8.85 \times 10^{-12})(0.40)}{5.0 \times 10^{-5}} = \boxed{1.8 \times 10^{-7}\,\text{F}} = 0.18\,\mu\text{F}. \tag{19.3.2}$$

(b) The field must not exceed the dielectric strength, $24 \times 10^6\ \text{V/m}$, so

$$V_{\max} = E_{\max}d = (24 \times 10^6)(5.0 \times 10^{-5}) = \boxed{1.2 \times 10^3\,\text{V}}. \tag{19.3.3}$$

(c) $Q_{\max} = CV_{\max} = (1.8 \times 10^{-7})(1.2 \times 10^3) = \boxed{2.2 \times 10^{-4}\,\text{C}}$. With air between the same foils, the capacitance would be 2.6 times smaller and the maximum voltage 8 times smaller. Rolled up into a compact cylinder, sheets like these make the familiar film capacitor.

Checkpoint 19.3

(a) Why can't a dielectric constant be less than 1? *(Hint: which way do the induced charges point?)* (b) A capacitor filled with a dielectric is disconnected from its battery, and the dielectric is then removed. Does the voltage increase or decrease? Does the stored energy increase or decrease? Where does the change in energy come from?

Answer

Answers to Checkpoints

Checkpoint 19.3. (a) The field of the capacitor polarizes the dielectric, and the induced surface charges produce a field pointing *opposite* to the applied field. The field inside is therefore reduced, never increased, so $E = E_0/\kappa$ requires $\kappa \geq 1$. ($\kappa = 1$ only for a vacuum, which cannot be polarized.) (b) With the battery disconnected, Q is fixed. Removing the dielectric reduces C by the factor κ, so $V = Q/C$ *increases* and $U = Q^2/2C$ *increases*. The extra energy comes from the work you do pulling the dielectric out: the capacitor's field attracts the dielectric inward, so you must pull against that attraction.

This page titled 19.3: Dielectrics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.4: Capacitors in Parallel and in Series

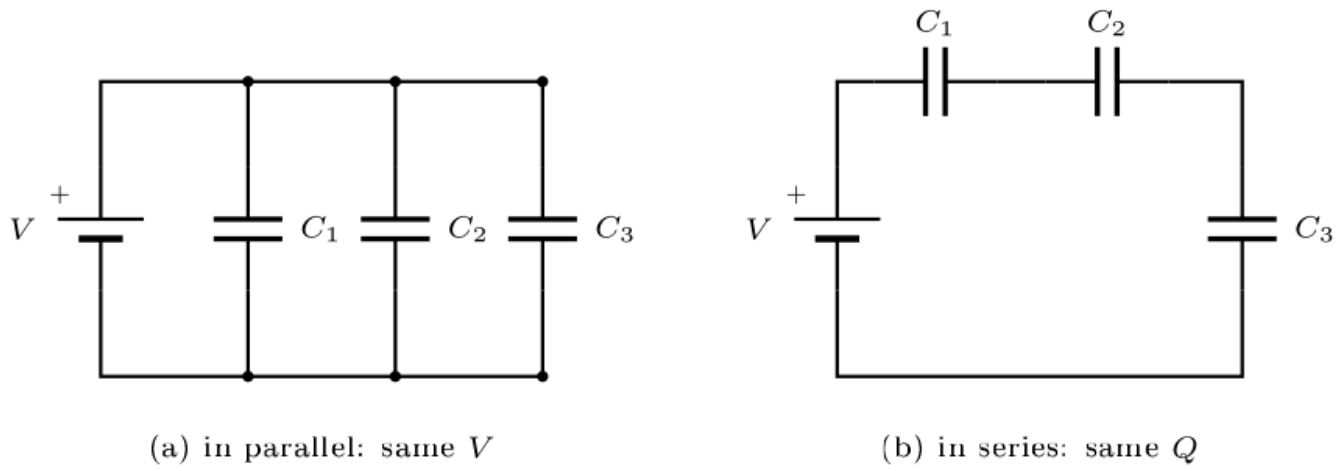


Figure 19.5. Capacitors in combination. (a) In parallel, every capacitor is connected across the same two wires and has the same potential difference; the charges add. (b) In series, the same charge sits on every capacitor; the potential differences add. In the battery symbol, the longer line marks the positive terminal.

•Figure description

Two circuit diagrams. In panel a, a battery of voltage V is connected to three capacitors C sub 1, C sub 2 and C sub 3 arranged side by side in parallel: each capacitor spans the same pair of horizontal wires, so each has the full voltage V across it. In panel b, the same battery drives three capacitors connected one after another in series around a single loop, so the same charge Q sits on each, and the individual voltages V sub 1, V sub 2 and V sub 3 add up to V.

Circuits often contain several capacitors connected together, and it is useful to replace a group of them with a single **equivalent capacitance** that stores the same charge at the same voltage. In a circuit diagram, a capacitor is drawn as two short parallel lines, like the plates, and a battery as a long line and a short line side by side, the long one being the positive terminal (Figure 19.5). Chapter 21 develops circuit diagrams and circuit analysis fully; here two simple rules are enough. The wires connecting the elements are conductors, so each wire and everything attached directly to it is at a single potential. And charge is conserved.

Capacitors in parallel. Capacitors connected side by side, so that one plate of each is attached to one wire and the other plate of each to another wire, are in **parallel** (Figure 19.5a). Each is connected directly across the same two wires, so each has the *same potential difference* V. Their charges are $Q_1 = C_1V$, $Q_2 = C_2V$, and so on, and the total charge supplied by the battery is

$$Q = Q_1 + Q_2 + \cdots = (C_1 + C_2 + \cdots)V. \tag{19.4.1}$$

A single capacitor that stores this total charge at the same voltage has

Key Equation: Capacitors in parallel

$$C_{\text{eq}} = C_1 + C_2 + C_3 + \cdots. \tag{19.9}$$

Capacitances in parallel simply add. This makes sense: connecting two parallel-plate capacitors in parallel is like making one capacitor with the combined plate area.

Capacitors in series. Capacitors connected one after another, so that the second plate of one is wired only to the first plate of the next, are in **series** (Figure 19.5b). The battery moves charge Q onto the first plate of the first capacitor and pulls Q off the last plate of the last. The inner plates are connected only to each other, and that isolated piece of metal starts out neutral and must stay neutral, so when $-Q$ appears on one of its plates, $+Q$ appears on the other. Every capacitor in series therefore carries the *same charge* Q. Their voltages, $V_1 = Q/C_1$, $V_2 = Q/C_2$, and so on, add up to the battery's voltage:

$$V = V_1 + V_2 + \cdots = Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \cdots\right). \tag{19.4.2}$$

A single capacitor that holds charge Q at voltage V has

Key Equation: Capacitors in series

$$\frac{1}{C_{\text{eq}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots. \tag{19.10}$$

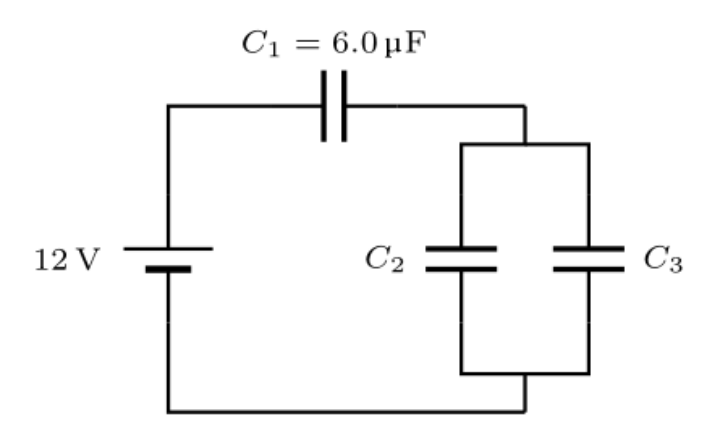
The equivalent capacitance of capacitors in series is always *smaller* than the smallest of them. Connecting two parallel-plate capacitors in series is like making one capacitor with a larger plate separation. For two capacitors in series, equation (19.10) can be written $C_{\text{eq}} = C_1C_2/(C_1 + C_2)$.

Note: A strategy for capacitor networks

Many networks can be reduced step by step. Find a group of capacitors that are purely in parallel or purely in series, replace it with its equivalent, redraw the circuit, and repeat until a single capacitor remains. Then work back out through the redrawn circuits. Remember which quantity is shared: *parallel* capacitors share the same V, and *series* capacitors share the same Q.

Example 19.8: A capacitor network

A $6.0\ \mu\text{F}$ capacitor C_1 is connected in series with a parallel pair, $C_2 = 2.0\ \mu\text{F}$ and $C_3 = 4.0\ \mu\text{F}$, and the combination is connected to a $12\ \text{V}$ battery. Find (a) the equivalent capacitance, (b) the charge on and voltage across each capacitor, and (c) the total stored energy.



•Figure description
A circuit diagram. A battery labeled 12 volts is connected in a loop. Along the top wire is capacitor C sub 1, 6 microfarads. On the right side, the loop splits into two parallel branches, one holding C sub 2, 2 microfarads, and the other holding C sub 3, 4 microfarads, which rejoin before returning along the bottom wire to the battery.

Solution

(a) C_2 and C_3 are in parallel: $C_{23} = 2.0 + 4.0 = 6.0\ \mu\text{F}$. This is in series with C_1:

$$C_{\text{eq}} = \frac{C_1C_{23}}{C_1 + C_{23}} = \frac{(6.0)(6.0)}{6.0 + 6.0} = \boxed{3.0\,\mu\text{F}}. \tag{19.4.3}$$

(b) The battery sees $3.0\ \mu\text{F}$, so the charge it supplies is $Q = C_{\text{eq}}V = (3.0\ \mu\text{F})(12\ \text{V}) = 36\ \mu\text{C}$. In series, C_1 and the parallel pair carry this same charge: $Q_1 = \boxed{36\,\mu\text{C}}$ and $V_1 = Q_1/C_1 = \boxed{6.0\,\text{V}}$. The pair gets the remaining $12 - 6.0 = 6.0\ \text{V}$, and each of its members has that full voltage: $V_2 = V_3 = \boxed{6.0\,\text{V}}$, so $Q_2 = C_2V_2 = \boxed{12\,\mu\text{C}}$ and $Q_3 = C_3V_3 = \boxed{24\,\mu\text{C}}$. As a check, $Q_2 + Q_3 = 36\ \mu\text{C} = Q_1$.

(c) $U = \frac{1}{2}C_{\text{eq}}V^2 = \frac{1}{2}(3.0 \times 10^{-6})(12)^2 = \boxed{216\,\mu\text{J}}$. Adding the individual energies, $\frac{1}{2}Q_1V_1 + \frac{1}{2}Q_2V_2 + \frac{1}{2}Q_3V_3 = 108 + 36 + 72 = 216$ μJ, gives the same total.

Checkpoint 19.4

You have two capacitors, $1\ \mu\text{F}$ and $3\ \mu\text{F}$. (a) Without calculating exactly, which connection, series or parallel, gives the larger equivalent capacitance? (b) Connected in series to a battery, which capacitor has the larger voltage across it? (c) Which stores more energy?

Answer

Answers to Checkpoints

Checkpoint 19.4. (a) Parallel. Parallel capacitances add ($1 + 3 = 4\ \mu\text{F}$), while in series the equivalent is smaller than the smaller one ($0.75\ \mu\text{F}$). (b) In series both carry the same charge Q, so $V = Q/C$ is larger across the *smaller* capacitor, the $1\ \mu\text{F}$ one: it has three times the voltage of the other. (c) Also the $1\ \mu\text{F}$ one. With equal charges, $U = Q^2/2C$ is larger for the smaller capacitance, again by a factor of 3.

This page titled 19.4: Capacitors in Parallel and in Series is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.S: Capacitance (Summary)

Summary

- A **capacitor** is two conductors carrying charges $\pm Q$. Its **capacitance** $C = Q/V$ depends only on geometry and on the material between the conductors, and is measured in farads ($1\ \text{F} = 1\ \text{C/V}$).
- Parallel plates: $C = \epsilon_0 A/d$. Cylindrical (coaxial): $C = 2\pi\epsilon_0 L/\ln(b/a)$. Spherical: $C = 4\pi\epsilon_0 ab/(b-a)$; an isolated sphere has $C = 4\pi\epsilon_0 R$.
- A charged capacitor stores energy $U = Q^2/2C = \frac{1}{2}CV^2 = \frac{1}{2}QV$. The energy resides in the field, with energy density $u_E = \frac{1}{2}\epsilon_0 E^2$.
- A **dielectric** of dielectric constant κ multiplies the capacitance, $C = \kappa C_0$, because its polarized molecules leave induced surface charges that reduce the field to E_0/κ. Its **dielectric strength** is the largest field it can withstand without breaking down.
- Capacitors in **parallel** share the same voltage, and $C_{\text{eq}} = C_1 + C_2 + \cdots$. Capacitors in **series** share the same charge, and $1/C_{\text{eq}} = 1/C_1 + 1/C_2 + \cdots$.

Key Equations

Capacitance

$$C = \frac{Q}{V} \tag{19.S.1}$$

Parallel-plate, cylindrical, and spherical capacitors

$$C = \frac{\epsilon_0 A}{d}, \qquad C = \frac{2\pi\epsilon_0 L}{\ln(b/a)}, \qquad C = 4\pi\epsilon_0 \frac{ab}{b-a} \tag{19.S.2}$$

Energy stored in a capacitor

$$U = \frac{Q^2}{2C} = \frac{1}{2}CV^2 = \frac{1}{2}QV \tag{19.S.3}$$

Energy density of an electric field

$$u_E = \frac{1}{2}\epsilon_0 E^2 \tag{19.S.4}$$

Capacitance with a dielectric

$$C = \kappa C_0, \qquad E = \frac{E_0}{\kappa}, \qquad \epsilon = \kappa\epsilon_0 \tag{19.S.5}$$

Capacitors in parallel and in series

$$C_{\text{eq}} = C_1 + C_2 + \cdots, \qquad \frac{1}{C_{\text{eq}}} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots \tag{19.S.6}$$

This page titled 19.S: Capacitance (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

19.E: Capacitance (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $\epsilon_0 = 8.85 \times 10^{-12}\ \text{F/m}$, $k_e = 8.99 \times 10^{9}\ \text{N}\cdot\text{m}^2/\text{C}^2$, and the values in Table 19.1.

Section 19.1 — Capacitance

19.1 ★ A capacitor holds $30\ \mu\text{C}$ when connected to a $9.0\ \text{V}$ battery. Find its capacitance. How much charge would it hold on a $1.5\ \text{V}$ battery?

19.2 ★ A parallel-plate capacitor has circular plates of radius $5.0\ \text{cm}$, $0.80\ \text{mm}$ apart in air. Find its capacitance, and the charge on it when it is connected to a $24\ \text{V}$ battery.

19.3 ★★ You want to build a $1.0\ \text{nF}$ air-filled parallel-plate capacitor with a plate separation of $0.10\ \text{mm}$. What plate area do you need? Would such a capacitor fit comfortably on a desk?

19.4 ★★ *[Conceptual.]* A charged parallel-plate capacitor is *disconnected* from its battery, and the plates are pushed closer together. Does each of the following increase, decrease, or stay the same: the capacitance, the charge, the potential difference, the field between the plates?

19.5 ★★ *[Real-world: electronics.]* A $20\ \text{m}$ length of coaxial cable has an inner conductor of diameter $0.90\ \text{mm}$ and an outer conductor of inner diameter $5.8\ \text{mm}$, with air between. Find its capacitance.

19.6 ★★ Two concentric conducting spheres have radii $8.0\ \text{cm}$ and $10.0\ \text{cm}$. Find the capacitance of this spherical capacitor. Compare it with the capacitance of a parallel-plate capacitor with the same separation and a plate area equal to the inner sphere's surface area, and explain why they differ by about 25% and in what limit they would agree.

19.7 ★★★ Show that when the gap $b - a$ between the conductors of a spherical capacitor is much smaller than a, equation (19.4) reduces to the parallel-plate result $C = \epsilon_0 A/d$ with $A = 4\pi a^2$ and $d = b - a$. Show the same for the cylindrical capacitor, using $\ln(1 + x) \approx x$ for small x.

Section 19.2 — Energy Stored in an Electric Field

19.8 ★ How much energy is stored in a $470\ \mu\text{F}$ capacitor charged to $25\ \text{V}$?

19.9 ★ What is the energy density in the field between the plates of a capacitor where the field is $5.0 \times 10^{5}\ \text{V/m}$?

19.10 ★★ *[Real-world: medicine.]* A defibrillator stores $360\ \text{J}$ in a $32\ \mu\text{F}$ capacitor. (a) To what voltage is it charged? (b) If the energy is delivered in $4.0\ \text{ms}$, what is the average power?

19.11 ★★ A $2.0\ \mu\text{F}$ capacitor charged to $100\ \text{V}$ is disconnected from its battery and connected to an uncharged $2.0\ \mu\text{F}$ capacitor. (a) Find the final charge on and voltage across each. (b) Find the total stored energy before and after. (c) Where did the missing energy go?

19.12 ★★ Use the energy density to find the energy stored in the field of an isolated conducting sphere of radius R carrying charge Q, by integrating u_E over all space outside the sphere. Check that your answer equals $Q^2/2C$ with $C = 4\pi\epsilon_0 R$.

19.13 ★★★ Show that the plates of a parallel-plate capacitor attract each other with a force $F = Q^2/(2\epsilon_0 A)$. *(Hint: with the battery disconnected, find how much the stored energy changes when the separation increases by dx; that change equals the work $F\,dx$ you must do.)* Explain why this is *not* QE, where E is the field between the plates.

Section 19.3 — Dielectrics

19.14 ★ A $12\ \text{pF}$ air capacitor is filled with mica. What is its new capacitance?

19.15 ★★ A parallel-plate capacitor with plates of area $0.020\ \mathrm{m}^2$ separated by $0.10\ \mathrm{mm}$ of Teflon is connected to a $60\ \mathrm{V}$ battery. Find (a) its capacitance, (b) the charge on the plates, (c) the field in the Teflon, and (d) the largest voltage the capacitor can withstand.

19.16 ★★ A capacitor charged to $200\ \mathrm{V}$ and then disconnected has a paper dielectric. The paper is carefully pulled out. What is the new voltage? By what factor does the stored energy change, and where does the extra energy come from?

19.17 ★★ *[Conceptual.]* Explain, in terms of polarization and induced charge, why the dielectric constant of water is so much larger than that of most plastics.

19.18 ★★★ A parallel-plate capacitor with plate area A and separation d is half filled by a dielectric slab of constant κ and thickness $d/2$, lying flat against one plate. Show that its capacitance is $C = 2\kappa\epsilon_0 A/\big((\kappa+1)d\big)$. *(Hint: treat it as two capacitors in series.)*

19.19 ★★★ A parallel-plate capacitor with plate area A and separation d has a dielectric slab of constant κ and full thickness d covering half the plate area. Find its capacitance. *(Hint: two capacitors in parallel.)* Compare with the previous problem for $\kappa = 3$.

Section 19.4 — Capacitors in Parallel and in Series

19.20 ★ Find the equivalent capacitance of a $3.0\ \mu\mathrm{F}$ and a $6.0\ \mu\mathrm{F}$ capacitor connected (a) in parallel and (b) in series.

19.21 ★★ The two capacitors of the previous problem are connected to a $9.0\ \mathrm{V}$ battery, first in parallel and then in series. For each connection find the charge on and voltage across each capacitor.

19.22 ★★ Three capacitors, $1.0\ \mu\mathrm{F}$, $2.0\ \mu\mathrm{F}$, and $3.0\ \mu\mathrm{F}$, are available. Find all the different equivalent capacitances that can be made using all three, and state which arrangement gives each.

19.23 ★★ *[Real-world: electronics.]* You need a capacitor rated for $600\ \mathrm{V}$, but you have only $10\ \mu\mathrm{F}$ capacitors rated for $200\ \mathrm{V}$. How can you combine them to get $10\ \mu\mathrm{F}$ rated for $600\ \mathrm{V}$? How many do you need?

19.24 ★★★ In the network of Example 19.8, capacitor C_2 fails and becomes a short circuit (a wire). Find the new charge on and voltage across C_1 and C_3, and the new total energy stored.

19.25 ★★★ *[Real-world: touchscreens.]* A capacitive touchscreen senses a finger because the finger changes the capacitance of a small pair of electrodes. Model one sensing element as a parallel-plate capacitor with plates $5.0\ \mathrm{mm}$ square, $0.50\ \mathrm{mm}$ apart, filled with glass of $\kappa = 7.0$. (a) Find its capacitance. (b) A finger adds a parallel path equivalent to a capacitance of about $1\ \mathrm{pF}$. By what percentage does the total capacitance change? (c) Why is a gloved finger usually not detected?

19.26 ★★★ *[Real-world: computer memory.]* A memory cell in a DRAM chip stores a bit as charge on a capacitor of about $25\ \mathrm{fF}$ ($25\times10^{-15}\ \mathrm{F}$) charged to $1.0\ \mathrm{V}$. (a) How many electrons represent a stored “1”? (b) The charge leaks away slowly, so each cell must be “refreshed” about every $64\ \mathrm{ms}$. If a cell loses 10% of its charge in that time, what average leakage current does that represent?

19.27 ★★★ Two identical parallel-plate capacitors, each of capacitance C, are connected in parallel to a battery of voltage V and then disconnected. A dielectric slab of constant κ is then slid into one of them. Find the new voltage across the pair, the charge on each, and the change in the total stored energy.

This page titled 19.E: Capacitance (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20: CURRENT AND RESISTANCE

Flip a light switch and the lamp across the room comes on at once. It is tempting to picture electrons racing from the switch to the bulb, but the truth is stranger. The electrons in the wire are moving at a crawl, a fraction of a millimeter per second, slower than a snail. An electron leaving the switch would take hours to reach the lamp. How can something so slow deliver energy so fast?

To this point, we've studied charges at rest, or *electrostatics*. In a conductor at rest the field inside is zero and nothing moves (Section 17.5). Starting with this chapter, we let the charges move. This is *electrodynamics*. A battery or generator maintains a field inside a wire, and the free electrons respond by flowing, forming an **electric current**. We will describe that current both on the scale of a whole wire and at the level of individual electrons, see why every material resists it to some degree, and find how much energy it delivers. By the end, the paradox of the slow electrons and the instant lamp will have a simple answer.

Learning Objectives

- Define electric current and current density, and relate them to the drift velocity of charge carriers.
- Define resistance, and calculate it from the resistivity and dimensions of a conductor, including its dependence on temperature.
- Apply Ohm's law, and distinguish ohmic from nonohmic devices using current–voltage graphs.
- Calculate the power delivered to a circuit element and dissipated in a resistor, and the cost of electrical energy.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

20.1: Electric Current
20.2: Resistance
20.3: Ohm's Law
20.4: Electrical Power
20.S: Current and Resistance (Summary)
20.E: Current and Resistance (Problems)

This page titled 20: Current and Resistance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.1: Electric Current

Any net flow of charge is an **electric current**. If a net charge dQ passes through a cross-section of a wire in a time dt, the current is the rate of flow,

Key Equation: Electric current

$$I = \frac{dQ}{dt}, \qquad I = \frac{\Delta Q}{\Delta t} \text{ (steady current)}. \tag{20.1}$$

The second form holds when the current is steady, so that the same charge passes each second. The SI unit of current is the **ampere** (A), usually shortened to *amp* in everyday speech. One ampere is one coulomb per second: $1\ \text{A} = 1\ \text{C/s}$. It is named for André-Marie Ampère, whose work on the magnetic effects of currents you will meet in Chapter 23. A phone charger delivers one or two amperes, a household circuit up to fifteen or twenty, and the starter motor of a car a couple of hundred. The signals in a computer chip involve microamperes or less.

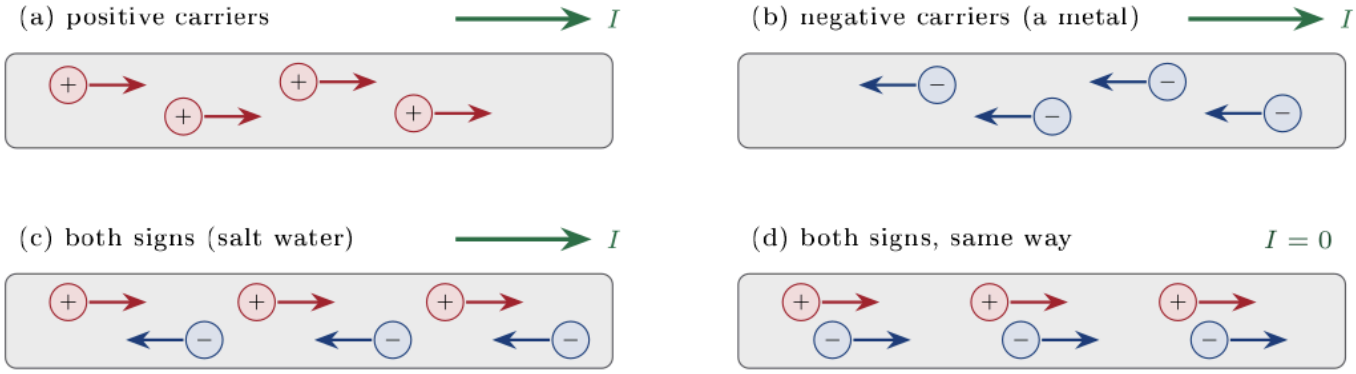


Figure 20.1. The direction of current. (a) Positive charges moving to the right, (b) negative charges moving to the left, and (c) both at once all carry net positive charge to the right, so in each case the current I points to the right. In a metal wire the carriers are electrons, as in (b), so they move opposite to the current. (d) When equal amounts of positive and negative charge move the same way, as when a neutral object is simply carried along, no net charge crosses and there is no current.

•Figure description

Four strips of conducting material arranged in a two-by-two grid. Panel a, top left: positive charges, drawn as red circles with plus signs, move to the right, and a green current arrow labeled I points to the right. Panel b, top right: negative charges, drawn as blue circles with minus signs, move to the left, as in a metal, and the current I still points to the right. Panel c, bottom left: positive and negative ions move in opposite directions at the same time, as in salt water, and the current I points to the right. Panel d, bottom right: equal numbers of positive and negative charges all move to the right together, as when a neutral object is carried along; the label says I equals zero, because no net charge crosses.

The direction of current. Current is a flow with a direction, but a charge can carry it either way: positive charge moving to the right and negative charge moving to the left both carry charge to the right (Figure 20.1). By convention, the direction of current is the direction in which *positive* charge would flow. This is the choice Chapter 15 warned about. In a metal wire, the charge carriers are electrons, so they actually move *opposite* to the direction of the current. In other conductors, such as salt water, the ions of both signs move, in opposite directions, and both contribute to a current in the same direction. For almost everything in circuits the distinction does not matter, and we will simply speak of current flowing from the high-potential end of a wire toward the low-potential end, the way positive charge would.

Current needs a field. In Chapter 17 we found that the field inside a conductor in electrostatic equilibrium is zero. A conductor carrying a current is *not* in equilibrium: charges are moving through it. To keep them moving there must be a field inside the conductor, pointing along the wire, and so a potential difference exists between its ends. That is the job of a battery or a generator: to maintain a potential difference, and with it a field, that keeps the charges flowing. The charge carriers are not used up. In a steady current, as much charge flows out of any section of wire as flows in, so the current is the same at every point along a single wire, and at a junction where wires meet, the total current flowing in equals the total flowing out. Both facts are consequences of

the conservation of charge, and both will be central to the circuits of Chapter 21.

Example 20.1: Charge through a phone charger

A phone charges for $1.5\ \text{h}$ at a steady current of $2.0\ \text{A}$. (a) How much charge passes through the charging cable? (b) How many electrons is that?

Solution

(a) With $\Delta t = (1.5)(3600) = 5400\ \text{s}$,

$$\Delta Q = I\,\Delta t = (2.0\,\text{A})(5400\,\text{s}) = \boxed{1.1 \times 10^4\,\text{C}}. \tag{20.1.1}$$

(b) Dividing by the charge of one electron,

$$N = \frac{\Delta Q}{e} = \frac{1.08 \times 10^4}{1.602 \times 10^{-19}} = \boxed{6.7 \times 10^{22}}\ \text{electrons}, \tag{20.1.2}$$

about a tenth of a mole. Battery capacities are often quoted in milliampere-hours (mAh), which is a unit of charge: $1\ \text{mAh} = (10^{-3}\ \text{A})(3600\ \text{s}) = 3.6\ \text{C}$. This charge corresponds to $3000\ \text{mAh}$, a typical phone battery.

Current Density

The current tells how much charge crosses a whole cross-section of a wire each second. To describe the flow at each point inside the conductor, we use the **current density** $\mathbf{J}$, the current per unit area through a surface perpendicular to the flow. For a current I spread uniformly over a cross-section of area A,

$$J = \frac{I}{A}, \tag{20.2}$$

measured in A/m^2. The current density is a vector, pointing in the direction of the current, that is, the direction positive charges move, and so along the field in the conductor.

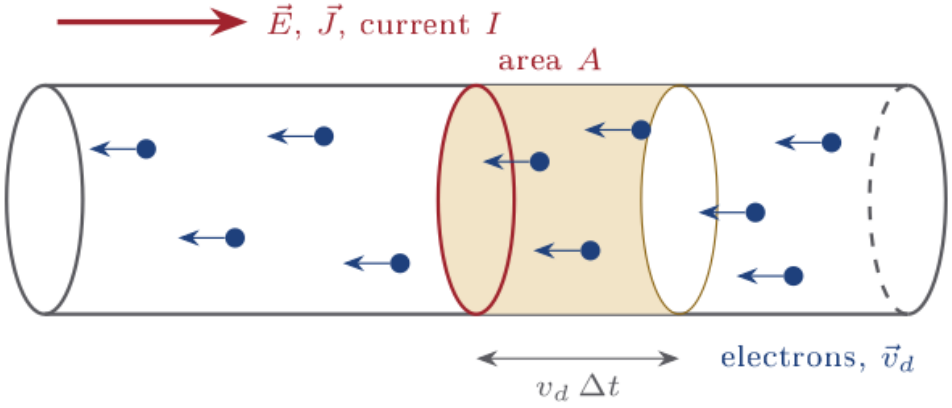


Figure 20.2. Drift of charge carriers in a wire. The electrons (blue) drift opposite to the field, with speed v_d. In a time Δt, all the carriers in the shaded slice of length $v_d\,\Delta t$ cross the red cross-section, carrying charge $nqAv_d\,\Delta t$. The current and the current density point along the field.

•Figure description

A section of a cylindrical wire of cross-sectional area A. The electric field E and the current density J point to the right along the wire. Small electrons inside drift to the left, opposite the field, with drift velocity v sub d. A shaded slice of the wire of length v sub d times delta t, just to the right of a marked cross-section, contains all the electrons that will cross that cross-section in time delta t.

Drift velocity. Current density connects directly to the motion of the charge carriers. Suppose a conductor contains n charge carriers per unit volume, each of charge q, and that they move along the wire with an average velocity v_d, called the **drift velocity**.

In a time Δt, every carrier within a distance $v_d\ \Delta t$ of a given cross-section passes through it (Figure 20.2). Those carriers fill a cylinder of volume $A v_d\ \Delta t$, so the charge that crosses is $\Delta Q = nqA v_d\ \Delta t$, and the current is $I = \Delta Q/\Delta t = nqA v_d$. Dividing by A:

Key Equation: Current density and drift velocity

$$\mathbf{J} = nq\mathbf{v}_d, \qquad I = nqAv_d. \tag{20.3}$$

For electrons, $q = -e$, so $\mathbf{v}_d$ points opposite to $\mathbf{J}$, as it must.

Example 20.2: The drift speed of electrons in a copper wire

A copper wire of the size used in household wiring has a diameter of $2.05\ \text{mm}$ and carries a current of $10.0\ \text{A}$. Copper has about one free electron per atom, which gives $n = 8.49 \times 10^{28}\ \text{m}^{-3}$. Find (a) the current density and (b) the drift speed of the electrons.

Solution

(a) The cross-sectional area is $A = \pi(1.025 \times 10^{-3})^2 = 3.30 \times 10^{-6}\ \text{m}^2$, so

$$J = \frac{I}{A} = \frac{10.0}{3.30 \times 10^{-6}} = \boxed{3.03 \times 10^{6}\,\text{A/m}^2}. \tag{20.1.3}$$

(b) From equation (20.3),

$$v_d = \frac{J}{ne} = \frac{3.03 \times 10^{6}}{(8.49 \times 10^{28})(1.602 \times 10^{-19})} = \boxed{2.2 \times 10^{-4}\,\text{m/s}}, \tag{20.1.4}$$

about a fifth of a millimeter per second. At this rate an electron needs over an hour to travel a meter of wire.

Why the lamp comes on at once. If the electrons crawl, why does the lamp light instantly? Because the wire is *already full* of free electrons, from the switch all the way to the filament. When the switch closes, the change in the field travels along the wire at nearly the speed of light, and electrons everywhere in the circuit, including those already in the filament, start drifting almost at once. It is like a garden hose that is already full of water: turn on the tap and water comes out of the far end immediately, even though no drop of water has traveled from the tap to the nozzle. The energy is carried by the field, not by the electrons' journey.

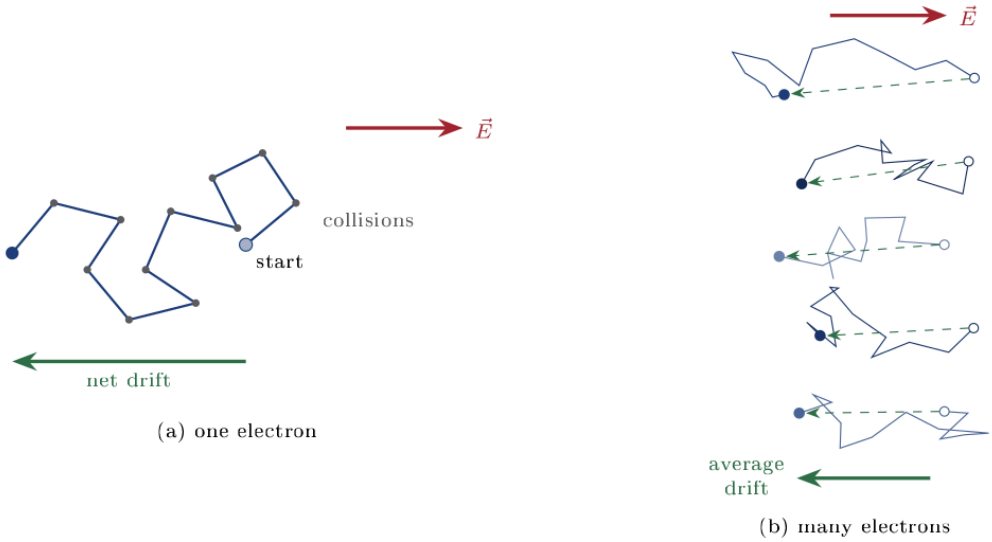


Figure 20.3. (a) The path of a conduction electron. Between collisions with the vibrating ions it moves fast in random directions. The field bends each short path slightly, opposite to $\mathbf{E}$ for a negative charge, so over many collisions the electron drifts slowly against the field. (b) Many electrons at once (open circles mark where each starts, filled circles where it ends). Each path looks random, but the net displacements (dashed) all lean against the field. Averaged over the enormous number of electrons in a wire, the random parts cancel and only this slow drift remains.

•Figure description

The zigzag path of a single electron in a metal, made of many short straight segments between collisions, pointing in random directions, with a gray dot at each collision. The field E points to the right. The path starts at a point marked start and, after many collisions, ends some distance to the left, opposite the field; an arrow below, labeled net drift, points to the left. Panel b shows five electrons at once, each starting at an open circle and following its own zigzag path through random collisions. Their paths look aimless, but every one ends up somewhat to the left of where it started, opposite the field E, shown by a dashed green arrow from each start to each end. A large green arrow labeled average drift points to the left.

The drift speed is also tiny compared with the electrons' random motion. Even with no current, the free electrons in a metal dart about at speeds of order $10^6\ \mathrm{m/s}$, colliding constantly with the vibrating ions of the metal. Their velocities point in every direction and average to zero. The field adds a slight bias to this frantic random motion, a slow drift superimposed on it (Figure 20.3a). Follow many electrons at once and the pattern is clearer: every path looks aimless, yet on average they all creep the same way (Figure 20.3b). The drift velocity is the average of all that motion, and it is billions of times smaller than the typical speed.

Checkpoint 20.1

Two copper wires carry the same current. Wire B has twice the diameter of wire A. (a) Compare the current densities in the two wires. (b) Compare the drift speeds. (c) A wire made of a metal with twice as many free electrons per cubic meter as copper carries the same current density as wire A. How does its drift speed compare?

Answer

Answers to Checkpoints

Checkpoint 20.1. (a) Doubling the diameter quadruples the area, so with the same current $J = I/A$ is four times smaller in wire B. (b) The drift speed is $v_d = J/(ne)$, and the two wires have the same n (both copper), so wire B's drift speed is also four times smaller. (c) With the same J but twice the carrier density, $v_d = J/(ne)$ is half that in wire A: more carriers share the work, so each can move more slowly.

This page titled 20.1: Electric Current is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.2: Resistance

Apply the same potential difference across two different objects, a length of copper wire and a length of nichrome wire, say, or a glass rod, and the currents that result are very different. The object's **resistance** measures how strongly it opposes the flow of charge. It is defined as the ratio of the potential difference across the object to the current through it:

Key Equation: Resistance

$$R = \frac{V}{I}. \tag{20.4}$$

The SI unit of resistance is the **ohm** (Ω), named for Georg Simon Ohm: $1\ \Omega = 1\ \mathrm{V/A}$. A copper connecting wire has a resistance of a small fraction of an ohm; the filament of a lamp a few hundred ohms; the human body, from hand to hand with dry skin, tens of thousands.

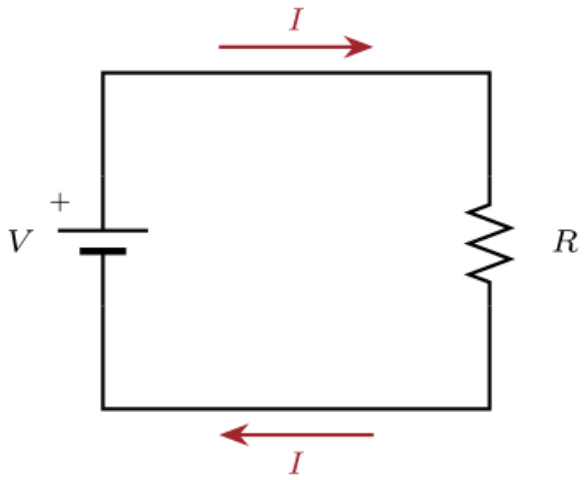


Figure 20.4. A battery connected to a resistor. The conventional current I leaves the battery's positive terminal (the longer line), passes through the resistor from its high-potential end to its low-potential end, and returns to the negative terminal.

•Figure description

A simple circuit diagram. A battery of voltage V, drawn with a long line for its positive terminal on top, is connected by wires around a rectangular loop to a resistor R, drawn as a zigzag, on the right side. Arrows on the wires show the conventional current I flowing out of the positive terminal, clockwise around the loop, down through the resistor and back into the negative terminal.

A component designed to have a particular resistance is a **resistor**, drawn in circuit diagrams as a zigzag line (Figure 20.4). Resistors are used to control currents and to set voltages throughout electronic circuits.

The analogy with water is useful. A potential difference is like a difference in water pressure between the two ends of a pipe; the current is like the rate at which water flows through it; and the resistance is like the narrowness and roughness of the pipe. A given pressure difference pushes more water through a wide, smooth pipe than through a narrow, rough one.

Example 20.3: The resistance of a lamp and of a person

(a) A lamp draws $0.50\ \mathrm{A}$ from a $120\ \mathrm{V}$ outlet. What is its resistance? (b) With dry skin, the resistance of a person's body from hand to hand is about $10^5\ \Omega$; with wet skin it can be as low as $1000\ \Omega$. What current would flow through the body if the hands touched the two terminals of a $120\ \mathrm{V}$ outlet in each case?

Solution

(a) $R = V/I = 120/0.50 = \boxed{240\,\Omega}$.

(b) With dry skin, $I = V/R = 120/10^5 = 1.2\ \text{mA}$, which is enough to feel as a tingle. With wet skin, $I = 120/1000 = \boxed{0.12\ \text{A}}$. A current of a tenth of an ampere passing through the chest can stop the heart. It is the current, not the voltage directly, that does the harm, and the body's resistance determines the current. This is why electrical appliances and water should never meet.

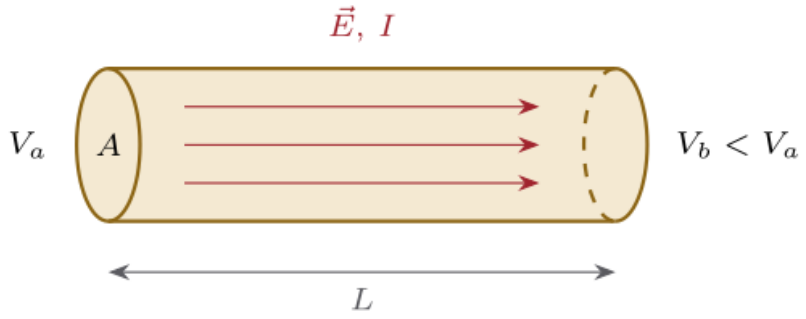


Figure 20.5. A uniform conductor of length L and cross-sectional area A. A potential difference $V = V_a - V_b$ drives a current I through it. Its resistance is $R = \rho L/A$.

•Figure description

A uniform cylindrical conductor of length L and cross-sectional area A, with a potential difference V applied between its ends. The left end is at the higher potential. A uniform electric field E points along the conductor from left to right, and a current I flows through it in the same direction.

Resistivity

The resistance of an object depends both on the material it is made of and on its shape and size. Experiment shows that a uniform wire or bar of length L and cross-sectional area A (Figure 20.5) has a resistance

Key Equation: Resistance of a uniform conductor

$$R = \rho \frac{L}{A}. \tag{20.5}$$

Doubling the length doubles the resistance: the charge must be pushed twice as far. Doubling the area halves it: there are twice as many parallel paths for the charge. The constant ρ, the **resistivity** of the material, depends only on the material and its temperature, not on the object's shape. Its unit is the ohm-meter ($\Omega \cdot$m). Its reciprocal, $\sigma = 1/\rho$, is the **conductivity**. (The same symbols ρ and σ also stand for charge densities; context makes the meaning clear.)

For a uniform conductor, the field inside is $E = V/L$ and the current density is $J = I/A$. Substituting both into $R = V/I = \rho L/A$ gives a relation between the field and the current density at each point in the material, the microscopic form of the same idea:

Key Equation: Field and current density in a material

$$\mathbf{E} = \rho \mathbf{J}, \qquad \text{or} \qquad \mathbf{J} = \sigma \mathbf{E}. \tag{20.6}$$

Table 20.1 lists the resistivities of some materials. Their range is extraordinary. The best conductor, silver, and a good insulator such as Teflon differ by some thirty orders of magnitude, a larger range than for almost any other physical property. Between them lie the **semiconductors**, such as silicon and germanium, whose resistivity can be adjusted over many orders of magnitude by adding tiny amounts of impurities.

Table 20.1. Approximate resistivities at $20\,^\circ\text{C}$, and temperature coefficients of resistivity near room temperature.

Material	Resistivity ρ ($\Omega \cdot$m)	Temperature coefficient α (K^{-1})
Silver	1.59×10^{-8}	3.8×10^{-3}

Material	Resistivity ρ ($\Omega \cdot$m)	Temperature coefficient α (K^{-1})
Copper	1.68×10^{-8}	3.9×10^{-3}
Gold	2.44×10^{-8}	3.4×10^{-3}
Aluminum	2.65×10^{-8}	3.9×10^{-3}
Tungsten	5.6×10^{-8}	4.5×10^{-3}
Iron	9.7×10^{-8}	5.0×10^{-3}
Platinum	1.06×10^{-7}	3.9×10^{-3}
Nichrome	1.10×10^{-6}	0.4×10^{-3}
Carbon (graphite)	3.5×10^{-5}	-0.5×10^{-3}
Germanium (pure)	0.46	-48×10^{-3}
Silicon (pure)	2.3×10^{3}	-75×10^{-3}
Glass	10^{10} to 10^{14}	—
Teflon	10^{22} to 10^{24}	—

Why metals resist. A simple model, first proposed by Paul Drude in 1900, explains where resistivity comes from. Between collisions with the metal's ions, a free electron is accelerated by the field, gaining a velocity $-eE\tau/m$ over the average time τ between collisions, the **mean free time**. Each collision scrambles its direction, erasing that extra velocity, and the process starts again. The drift velocity is the average gain, $v_d = eE\tau/m$ in size, so the current density is $J = nev_d = (ne^2\tau/m)E$. Comparing with $J = E/\rho$,

$$\rho = \frac{m}{ne^2\tau}. \tag{20.7}$$

A material resists current more strongly if it has fewer carriers (n small) or if they collide more often (τ short).

Example 20.4: Resistance of a copper wire

Find the resistance of a $10.0\ \mathrm{m}$ length of copper wire of diameter $1.00\ \mathrm{mm}$. What is the potential difference across it when it carries $2.0\ \mathrm{A}$?

Solution

The area is $A = \pi(0.500 \times 10^{-3})^2 = 7.85 \times 10^{-7}\ \mathrm{m}^2$, so

$$R = \rho\frac{L}{A} = (1.68 \times 10^{-8})\frac{10.0}{7.85 \times 10^{-7}} = \boxed{0.214\,\Omega}. \tag{20.2.1}$$

The potential difference is $V = IR = (2.0)(0.214) = \boxed{0.43\,\mathrm{V}}$. Ten meters of ordinary wire has only a fifth of an ohm of resistance, which is why the connecting wires in a circuit can usually be treated as having none at all.

Example 20.5: The mean free time in copper

Use equation (20.7) to find the mean free time between collisions for the conduction electrons in copper.

Solution

Solving for τ, with $n = 8.49 \times 10^{28}\ \mathrm{m}^{-3}$,

$$\tau = \frac{m}{ne^2\rho} = \frac{9.11 \times 10^{-31}}{(8.49 \times 10^{28})(1.602 \times 10^{-19})^2(1.68 \times 10^{-8})} = \boxed{2.5 \times 10^{-14}\,\mathrm{s}}. \tag{20.2.2}$$

An electron in copper suffers some forty trillion collisions every second. The simple model gives the right size for τ, but explaining its value properly, and why electrons pass freely through a perfect crystal and scatter only from its imperfections and vibrations, requires the quantum mechanics of Part III.

Resistivity and temperature. As a metal warms, its ions vibrate more vigorously, electrons collide with them more often, τ shortens, and the resistivity rises. Over moderate temperature ranges the increase is nearly linear:

Key Equation: Temperature dependence of resistivity

$$\rho = \rho_0 \left[1 + \alpha(T - T_0)\right], \tag{20.8}$$

where ρ_0 is the resistivity at a reference temperature T_0 (usually $20\,^\circ\mathrm{C}$) and α is the **temperature coefficient of resistivity**. The resistance of a given conductor changes in the same proportion, $R = R_0[1 + \alpha(T - T_0)]$. For semiconductors α is *negative*: heating frees more charge carriers, and the increase in n outweighs the decrease in τ.

Example 20.6: A platinum resistance thermometer

A platinum thermometer has a resistance of $100.0\ \Omega$ at $0\,^\circ\mathrm{C}$. Its resistance in a pot of liquid is $138.5\ \Omega$. What is the temperature of the liquid? (For platinum near these temperatures, $\alpha = 3.85 \times 10^{-3}\ \mathrm{K}^{-1}$ relative to $0\,^\circ\mathrm{C}$.)

Solution

Solving $R = R_0[1 + \alpha(T - T_0)]$ for T,

$$T = T_0 + \frac{R/R_0 - 1}{\alpha} = 0 + \frac{1.385 - 1}{3.85 \times 10^{-3}} = \boxed{100\,^\circ\mathrm{C}}. \tag{20.2.3}$$

The liquid is boiling water. Platinum resistance thermometers of exactly this kind, called "Pt100" sensors, are among the most accurate and widely used thermometers in industry.

Note: Superconductors

In 1911 Heike Kamerlingh Onnes cooled mercury in liquid helium and found that below $4.2\ \mathrm{K}$ its resistance did not merely become small: it vanished completely. A current set up in a loop of **superconductor** circulates for years with no battery at all. Many metals and compounds become superconducting when cold enough, some ceramics at temperatures above $77\ \mathrm{K}$, the boiling point of liquid nitrogen. Superconducting wire carries the enormous currents in the magnets of MRI scanners and particle accelerators. Superconductivity cannot be explained by the classical Drude picture at all; it is a quantum effect, in which electrons pair up and move through the crystal together without scattering.

Checkpoint 20.2

(a) A wire is stretched to twice its original length, keeping its volume fixed. By what factor does its resistance change? *(Hint: what happens to its cross-sectional area?)* (b) Why does an incandescent bulb draw its largest current in the instant after it is switched on, and why do bulbs most often burn out at that moment?

Answer

Answers to Checkpoints

Checkpoint 20.2. (a) Doubling the length at fixed volume halves the cross-sectional area. Since $R = \rho L/A$, the length doubling doubles R and the area halving doubles it again: the resistance rises by a factor of 4. (b) A metal's resistivity rises with temperature. A cold filament has a resistance several times smaller than when it glows at operating temperature, so at switch-on it draws a surge of current, several times its normal value, until it heats up. That surge heats the filament fastest at its thinnest spot, and the sudden thermal stress is what most often breaks a worn filament at that moment.

This page titled 20.2: Resistance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.3: Ohm's Law

Equation (20.4) defines the resistance of any object, $R = V/I$. By itself it says nothing about how R behaves. For many materials, however, and in particular for metals at constant temperature, experiment shows that the current is directly proportional to the applied potential difference over a very wide range. The resistance is then a constant, independent of V and I. This experimental fact is called **Ohm's law**, after Georg Ohm, who reported it in 1827:

Key Equation: Ohm's law

$$V = IR, \qquad R \text{ constant}. \tag{20.9}$$

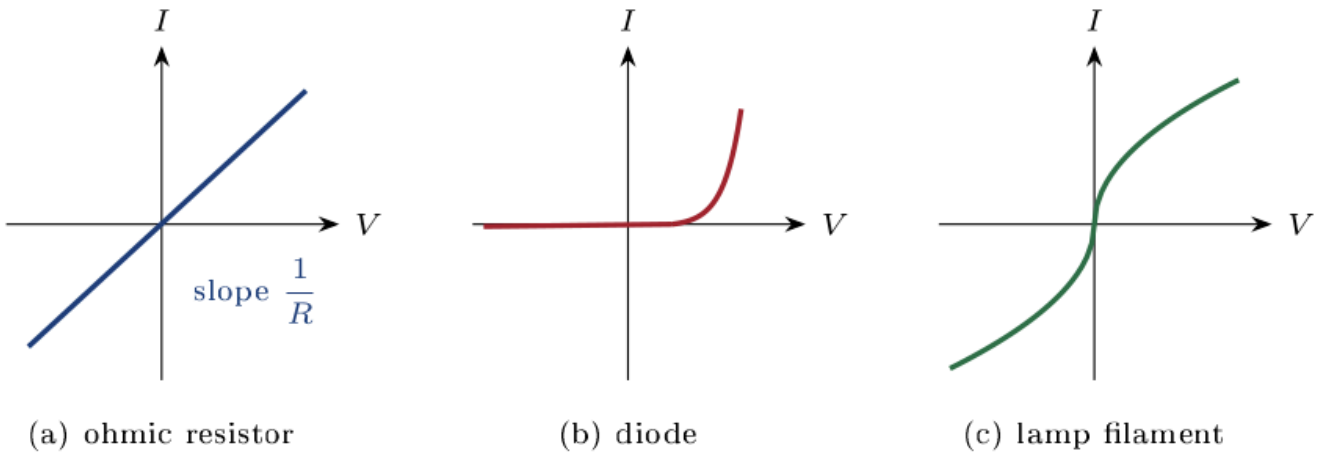


Figure 20.6. Current–voltage graphs. (a) An ohmic resistor: current proportional to voltage, a straight line of slope $1/R$. (b) A diode: almost no current in one direction, and a steeply rising current in the other once the voltage exceeds a threshold. (c) A lamp filament: its resistance rises as it heats, so the graph bends toward the V axis.

•Figure description

Three graphs of current I against potential difference V. Panel a, an ohmic resistor: a straight line through the origin, extending into both the positive and negative quadrants, with slope one over R. Panel b, a diode: the current is essentially zero for negative V and for small positive V, then rises very steeply beyond about 0.6 volts. Panel c, a lamp filament: the curve passes through the origin and rises steeply at first, then bends over and becomes less steep at larger V, because the hot filament's resistance is higher.

A material or device that obeys Ohm's law is called **ohmic**. Its graph of current against voltage is a straight line through the origin, with slope $1/R$ (Figure 20.6a). Ohm's law is not a fundamental law of nature like Coulomb's law or Gauss's law. It is an empirical rule that many materials obey, just as many springs obey Hooke's law. Many important devices do not obey it at all. A **diode**, for instance, conducts easily in one direction and hardly at all in the other (Figure 20.6b), which makes it the basic component for turning alternating current into direct current. Even a lamp filament is nonohmic in practice: as the current rises, the filament heats up and its resistance increases, so its graph bends over (Figure 20.6c).

For a nonohmic device, $R = V/I$ still defines a resistance at each operating point, but it changes with the voltage, and quoting a single "resistance" for the device is not meaningful. From here on, unless stated otherwise, resistors are ohmic, and $V = IR$ with a constant R.

Example 20.7: Using Ohm's law

A $470\ \Omega$ resistor is connected across a $9.0\ \text{V}$ battery. (a) What current flows through it? (b) What resistance would make the current $5.0\ \text{mA}$?

Solution

(a) $I = V/R = 9.0/470 = \boxed{19\,\text{mA}}$.

(b) $R = V/I = 9.0/(5.0 \times 10^{-3}) = \boxed{1.8\,\text{k}\Omega}$. Resistors are made in a standard series of values, and $1.8\ \text{k}\Omega$ is one of them.

Example 20.8: Reading a current–voltage graph

For the lamp of Figure 20.6c, suppose the current is $0.10\ \text{A}$ at $1.0\ \text{V}$ and $0.30\ \text{A}$ at $9.0\ \text{V}$. Find the filament's resistance at each point, and explain the difference.

Solution

At $1.0\ \text{V}$, $R = V/I = 1.0/0.10 = \boxed{10\,\Omega}$. At $9.0\ \text{V}$, $R = 9.0/0.30 = \boxed{30\,\Omega}$. Tripling the current raised the resistance threefold, because the filament is much hotter: with twenty-seven times the power going into it ($2.7\ \text{W}$ against $0.10\ \text{W}$) (Section 20.4), its temperature and hence its resistivity rise substantially. Note that R is V/I at each point, the slope of the line from the origin, not the slope of the curve.

Checkpoint 20.3

(a) A student says: "Ohm's law, $V = IR$, is just the definition of resistance, so everything obeys it." What is wrong with this statement? (b) Sketch the I–V graph of a resistor whose resistance *decreases* as the current increases.

Answer

Answers to Checkpoints

Checkpoint 20.3. (a) $R = V/I$ *defines* resistance for any device at any operating point, but Ohm's law is the stronger, empirical claim that R stays *constant*, that I is proportional to V. Many devices violate it: a diode's current is nearly zero one way and large the other, and a lamp's resistance changes as it heats. Ohm's law describes a class of materials, not a law of nature. (b) The graph passes through the origin and curves *upward*, growing steeper as V increases. The slope of an I–V graph is $1/R$, so a decreasing resistance means an increasing slope.

This page titled 20.3: Ohm's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.4: Electrical Power

A battery lifts charge to a higher potential, and a resistor lets it fall again. As charge dQ passes through a circuit element across which the potential drops by V, its potential energy falls by $dQ\ V$ (Section 18.1). In a resistor, that energy does not become kinetic energy of the carriers: their drift speed stays the same all along the wire. Instead, it goes into the collisions with the ions, making them vibrate more and heating the material. The rate at which the energy is delivered is

$$P = \frac{dU}{dt} = \frac{dQ}{dt}V, \tag{20.4.1}$$

and so the **electrical power** delivered to any circuit element is

Key Equation: Electrical power

$$P = IV. \tag{20.10}$$

This holds for any device, whether a resistor, a motor, a charging battery, or a lamp: current times potential difference is power, in watts, since $(1\ \text{C/s})(1\ \text{J/C}) = 1\ \text{J/s} = 1\ \text{W}$. For a resistor, using $V = IR$, the power dissipated as heat can be written in two more ways:

Key Equation: Power dissipated in a resistor

$$P = I^2R = \frac{V^2}{R}. \tag{20.11}$$

This heating is called **Joule heating**. It is exactly what a toaster, an electric kettle, or a space heater is designed to do, and it is an unwanted loss in wires, transformers, and electronic circuits, which is why computers need fans.

Electric utilities bill for energy, not power, in **kilowatt-hours** (kWh): the energy delivered in one hour at a power of one kilowatt, $1\ \text{kWh} = (1000\ \text{W})(3600\ \text{s}) = 3.6 \times 10^6\ \text{J}$.

Example 20.9: A space heater

A space heater is rated at $1500\ \text{W}$ on a $120\ \text{V}$ circuit. (a) What current does it draw, and what is its resistance when hot? (b) If it runs $6.0\ \text{h}$ a day for a 30-day month, and electrical energy costs $\$0.30$ per kWh, what does it cost to run?

Solution

(a) $I = P/V = 1500/120 = \boxed{12.5\,\text{A}}$, and $R = V/I = 120/12.5 = \boxed{9.6\,\Omega}$ (or $R = V^2/P$). That is most of the capacity of a typical $15\ \text{A}$ household circuit, which is why running a heater and a hair dryer on the same circuit trips the breaker.

(b) The energy used is $(1.5\ \text{kW})(6.0\ \text{h/day})(30\ \text{days}) = 270\ \text{kWh}$, costing $(270)(\$0.30) = \boxed{\$81}$.

Example 20.10: Why power lines use high voltage

A power plant sends $100\ \text{MW}$ to a city through transmission lines with a total resistance of $5.0\ \Omega$. Find the power lost as heat in the lines if the power is transmitted at (a) $10\ \text{kV}$ and (b) $500\ \text{kV}$.

Solution

The current in the lines is $I = P/V$, and the power lost in them is $P_{\text{loss}} = I^2R$.

(a) At $10\ \text{kV}$, $I = (100 \times 10^6)/(10 \times 10^3) = 1.0 \times 10^4\ \text{A}$, and

$$P_{\text{loss}} = (1.0 \times 10^4)^2(5.0) = \boxed{5.0 \times 10^8\ \text{W}}, \tag{20.4.2}$$

five times the power being sent. This is impossible in practice: the lines would melt long before delivering anything.

(b) At $500\ \text{kV}$, $I = 200\ \text{A}$, and

$$P_{\text{loss}} = (200)^2(5.0) = \boxed{2.0 \times 10^5\ \text{W}}, \tag{20.4.3}$$

only 0.2% of the power sent. Because the loss goes as I^2, raising the voltage fifty times cuts the loss by a factor of 2500. This is why long-distance lines run at hundreds of thousands of volts, and why the transformers of Chapter 25 are needed to step the voltage up for transmission and back down for use in the home.

Note: $P = I^2R$ or $P = V^2/R$?

The two formulas look contradictory: one says power grows with resistance, the other that it falls. Both are right; they answer different questions. For resistors carrying the *same current*, as in the transmission line, more resistance means more heating, $P = I^2R$. For resistors connected to the *same voltage*, as for appliances plugged into the same outlet, less resistance means more current and more heating, $P = V^2/R$. A $100\ \text{W}$ bulb has *less* resistance than a $60\ \text{W}$ bulb.

Checkpoint 20.4

A $60\ \text{W}$ and a $100\ \text{W}$ bulb are both designed for $120\ \text{V}$. (a) Which has the greater resistance? (b) Which draws more current when plugged in normally? (c) What happens to the power of the $100\ \text{W}$ bulb if it is run on $60\ \text{V}$, assuming its resistance stays the same?

Answer

Answers to Checkpoints

Checkpoint 20.4. (a) From $P = V^2/R$, $R = V^2/P$, so the lower-power bulb has the *greater* resistance: $(120)^2/60 = 240\ \Omega$ versus $(120)^2/100 = 144\ \Omega$. (b) The $100\ \text{W}$ bulb, since $I = P/V$: $0.83\ \text{A}$ versus $0.50\ \text{A}$. (c) At fixed resistance, $P = V^2/R$ scales as V^2, so halving the voltage cuts the power to a quarter: $25\ \text{W}$. (In reality the cooler filament's resistance would drop somewhat, so the bulb would draw a little more than this.)

This page titled 20.4: Electrical Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.S: Current and Resistance (Summary)

Summary

- **Electric current** is the rate of flow of charge, $I = dQ/dt$, measured in amperes ($1\ \text{A} = 1\ \text{C/s}$). Its direction is that of positive-charge flow; in metals the electrons move the opposite way. A steady current is the same all along a wire, and requires a field inside the conductor.
- **Current density** $\mathbf{J}$ is current per unit area. It is related to the **drift velocity** of the carriers by $\mathbf{J} = nq\mathbf{v}_d$, so $I = nqAv_d$. Drift speeds in wires are tiny, of order $0.1\ \text{mm/s}$; the field, and the energy, travel at nearly the speed of light.
- **Resistance** is $R = V/I$, in ohms. A uniform conductor has $R = \rho L/A$, where the **resistivity** ρ is a property of the material; microscopically $\mathbf{E} = \rho\mathbf{J}$, and the Drude model gives $\rho = m/(ne^2\tau)$. For metals, ρ rises with temperature, $\rho = \rho_0[1 + \alpha(T - T_0)]$; for semiconductors it falls.
- **Ohm's law**: for ohmic materials, $V = IR$ with R constant, so the I–V graph is a straight line through the origin. It is an empirical rule, not a law of nature; diodes and hot filaments are nonohmic.
- The **power** delivered to any circuit element is $P = IV$. In a resistor it becomes heat, $P = I^2R = V^2/R$. Electrical energy is billed in kilowatt-hours, $1\ \text{kWh} = 3.6 \times 10^6\ \text{J}$.

Key Equations

Electric current

$$I = \frac{dQ}{dt} \tag{20.S.1}$$

Current density and drift velocity

$$J = \frac{I}{A}, \qquad \mathbf{J} = nq\mathbf{v}_d, \qquad I = nqAv_d \tag{20.S.2}$$

Resistance, resistivity, and conductivity

$$R = \frac{V}{I}, \qquad R = \rho\frac{L}{A}, \qquad \mathbf{E} = \rho\mathbf{J}, \qquad \sigma = \frac{1}{\rho} \tag{20.S.3}$$

Temperature dependence of resistivity

$$\rho = \rho_0\left[1 + \alpha(T - T_0)\right] \tag{20.S.4}$$

Ohm's law

$$V = IR \quad (R \text{ constant}) \tag{20.S.5}$$

Electrical power

$$P = IV, \qquad P = I^2R = \frac{V^2}{R} \ \ (\text{resistor}) \tag{20.S.6}$$

This page titled 20.S: Current and Resistance (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

20.E: Current and Resistance (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $e = 1.602 \times 10^{-19}\ \text{C}$, $m_e = 9.11 \times 10^{-31}\ \text{kg}$, the resistivities of Table 20.1, and, for copper, $n = 8.49 \times 10^{28}\ \text{m}^{-3}$.

Section 20.1 — Electric Current

20.1 ★ A current of $3.5\ \text{A}$ flows in a wire for $4.0\ \text{min}$. How much charge passes any point in the wire, and how many electrons is that?

20.2 ★ *[Real-world: batteries.]* A phone battery is rated at $4500\ \text{mAh}$. (a) How many coulombs is that? (b) How long could it supply a steady $300\ \text{mA}$?

20.3 ★★ *[Conceptual.]* In a wire carrying a steady current, electrons move to the left. (a) In which direction does the current flow? (b) In which direction does the electric field inside the wire point? (c) Which end of the wire is at the higher potential?

20.4 ★★ *[Real-world: atmospheric physics.]* A lightning stroke transfers about $5.0\ \text{C}$ in about $0.20\ \text{ms}$. What is the average current?

Section 20.1.1 — Current Density

20.5 ★ A wire of diameter $1.6\ \text{mm}$ carries $5.0\ \text{A}$. Find the current density.

20.6 ★★ An aluminum wire carrying $8.0\ \text{A}$ has a cross-sectional area of $4.0\ \text{mm}^2$. Aluminum contributes about three free electrons per atom, giving $n = 1.8 \times 10^{29}\ \text{m}^{-3}$. Find the drift speed.

20.7 ★★ How long does it take a single electron to drift $2.0\ \text{m}$ along the wire of Example 20.2? Compare with the time for the signal to travel the same distance at nearly the speed of light.

20.8 ★★★ In a particle accelerator, a beam of protons moving at $3.0 \times 10^{7}\ \text{m/s}$ carries a current of $2.0\ \text{mA}$ and has a circular cross-section of radius $1.0\ \text{mm}$. Find (a) the current density and (b) the number of protons per cubic meter in the beam.

Section 20.2 — Resistance

20.9 ★ A $12\ \text{V}$ car headlight draws $4.5\ \text{A}$. What is its resistance?

20.10 ★★ *[Conceptual.]* Using the water-pipe analogy, explain why a long, thin wire has more resistance than a short, thick one of the same material.

Section 20.2.1 — Resistivity

20.11 ★ Find the resistance of a $50\ \text{m}$ aluminum wire of diameter $2.0\ \text{mm}$.

20.12 ★★ A nichrome heating element is to have a resistance of $24\ \Omega$ using wire of diameter $0.50\ \text{mm}$. How long must the wire be?

20.13 ★★ A copper wire and an aluminum wire have the same length and the same resistance. What is the ratio of their diameters? Which is lighter? (Densities: copper $8.96 \times 10^{3}\ \text{kg/m}^3$, aluminum $2.70 \times 10^{3}\ \text{kg/m}^3$.) This is why overhead power lines are made of aluminum.

20.14 ★★ A tungsten lamp filament has a resistance of $20\ \Omega$ at $20\,^{\circ}\text{C}$. Estimate its resistance at its operating temperature of $2500\,^{\circ}\text{C}$, using equation (20.8). (The linear formula is only rough over so large a range.)

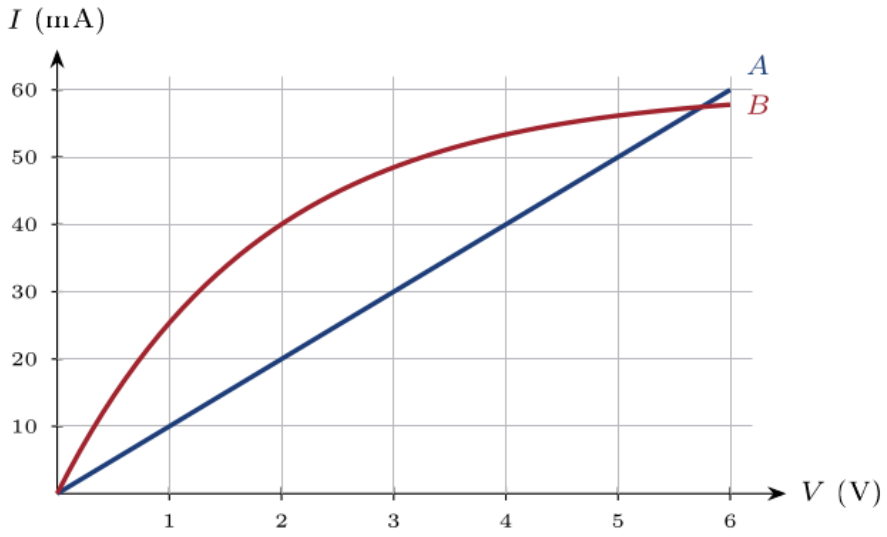
20.15 ★★ A copper wire has a resistance of $5.00\ \Omega$ at $20\,^\circ\mathrm{C}$. At what temperature is its resistance $6.00\ \Omega$?

20.16 ★★★ A hollow copper cylinder of length L has inner radius a and outer radius b. Current flows *radially*, from the inner surface to the outer. Show that the resistance is $R = (\rho/2\pi L)\ln(b/a)$. *(Hint: add up thin cylindrical shells in series.)*

Section 20.3 — Ohm's Law

20.17 ★ What current flows through a $2.2\ \mathrm{k\Omega}$ resistor connected across $5.0\ \mathrm{V}$?

20.18 ★★ *[Graph-reading.]* The graph shows current against voltage for two devices, A and B.



•Figure description
A graph of current I in milliamperes, from 0 to 60, against voltage V in volts, from 0 to 6. Device A is a straight line through the origin reaching 60 milliamperes at 6 volts. Device B is a curve through the origin that rises steeply at first, reaching about 40 milliamperes at 2 volts, then flattens, reaching about 58 milliamperes at 6 volts.

a. Which device is ohmic? Find its resistance.
b. Find the resistance of B at $1.0\ \mathrm{V}$ and at $5.0\ \mathrm{V}$.
c. Does the resistance of B increase or decrease with voltage? What kind of device might behave this way?

20.19 ★★ *[Conceptual.]* A battery is connected to a length of copper wire and a current flows. The wire is then cut in half and one half is connected to the same battery. Assuming the battery's voltage stays fixed, what happens to the current?

Section 20.4 — Electrical Power

20.20 ★ A $9.0\ \mathrm{V}$ battery supplies $25\ \mathrm{mA}$ to a circuit. What power does it deliver?

20.21 ★ What is the resistance of a $1200\ \mathrm{W}$ hair dryer designed for $120\ \mathrm{V}$, and what current does it draw?

20.22 ★★ *[Real-world: household energy.]* A refrigerator uses about $450\ \mathrm{kWh}$ per year. What is its average power? At $\$0.30$ per kWh, what does it cost to run for a year?

20.23 ★★ An electric kettle heats $1.0\ \mathrm{L}$ of water from $20\,^\circ\mathrm{C}$ to $100\,^\circ\mathrm{C}$ in $3.0\ \mathrm{min}$. Assuming all the electrical energy goes into the water (specific heat $4186\ \mathrm{J/(kg\cdot K)}$, Chapter 37), what power does it draw, and what current if it runs on $230\ \mathrm{V}$?

20.24 ★★ The same wire carries a current I in two circuits, and in the second the current is doubled. By what factor does the power dissipated in the wire change? If instead the voltage across a resistor is doubled, by what factor does its power change?

20.25 ★★★ In Example 20.10, what fraction of the power is lost if it is transmitted at $100\ \mathrm{kV}$? At what transmission voltage

would the loss be exactly 1%?

20.26 ★★★ *[Real-world: electrical safety.]* Household wiring uses copper wire of diameter $2.05\ \text{mm}$ on a $15\ \text{A}$ circuit. (a) Find the resistance per meter of the wire. (b) Find the power dissipated per meter at $15\ \text{A}$. (c) Explain why using a thinner wire, or overloading the circuit, can start a fire inside a wall, and what the circuit breaker is for.

20.27 ★★★ Show that the power dissipated per unit volume in a resistor is $J^2\rho = \sigma E^2$. Use it to find the power per cubic meter dissipated in the copper wire of Example 20.2.

20.28 ★★★ A cylindrical wire of radius R and length L has a resistivity that increases with distance from its axis, $\rho(r) = \rho_0(1 + r^2/R^2)$. A potential difference V is applied between its ends. (a) Find the current density as a function of r. (b) Find the total current, and hence the resistance.

This page titled 20.E: Current and Resistance (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21: DC CIRCUITS

A flashlight has a battery, a switch, and a bulb. A laptop has billions of transistors, thousands of resistors and capacitors, and a battery pack. The two could hardly look more different, yet they obey the same two rules. Charge is conserved, so every bit of current that flows into a junction flows out again. And energy is conserved, so a charge that travels once around any closed path through a circuit returns to the potential it started from. These two statements, Kirchhoff's laws, are all it takes to analyze any circuit powered by a steady source.

Chapter 20 studied a single resistor with a potential difference across it. This chapter builds circuits: it introduces the battery as a source of energy, the diagrams that engineers use to describe circuits, and the rules for combining resistors. It ends with circuits containing capacitors, where for the first time the current changes with time, rising or dying away on a time scale set by the product RC. Such circuits set the rhythm of blinking lights, the delay of a windshield wiper, and the timing of a heart pacemaker. All of the currents here are *direct currents* (dc), flowing in one direction; alternating currents come in Chapter 26.

Learning Objectives

- Read and draw circuit diagrams, and describe what the emf of a battery means.
- Apply Kirchhoff's junction and loop rules to single-loop and multiloop circuits, and account for the energy and power in a circuit.
- Find the equivalent resistance of resistors in series, in parallel, and in combination.
- Model a real battery as an ideal emf with an internal resistance, and find its terminal voltage and the power it delivers.
- Describe the charging and discharging of a capacitor through a resistor, using the time constant RC.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

21.1: Circuit Elements and Diagrams
21.2: Kirchhoff's Laws and the Basic Circuit
21.3: Multiple Resistors in a Circuit: Series vs. Parallel
21.4: Ideal vs. Real Batteries
21.5: RC Circuits
21.S: DC Circuits (Summary)
21.E: DC Circuits (Problems)

This page titled 21: DC Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.1: Circuit Elements and Diagrams

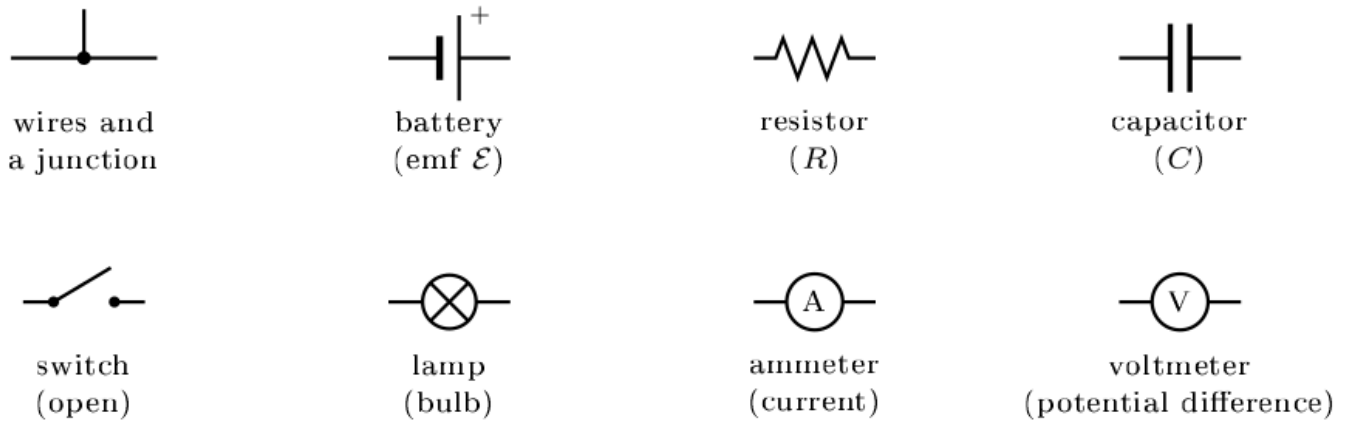


Figure 21.1. Symbols used in circuit diagrams. The long line of the battery symbol is its positive terminal.

•Figure description

The standard symbols used in circuit diagrams, in two rows of four. Top row: wires, drawn as straight lines, with a dot marking a junction where three wires meet; a battery, drawn as a long thin line and a short thick line side by side, the long line marked plus; a resistor, drawn as a zigzag; and a capacitor, drawn as two short parallel lines of equal length. Bottom row: an open switch, drawn as a gap with a hinged blade tilted up; a lamp, drawn as a circle with an X in it; an ammeter, drawn as a circle containing the letter A; and a voltmeter, drawn as a circle containing the letter V.

An **electric circuit** is a closed path, or a network of paths, around which charge can flow. It is built from a small number of kinds of element, joined by wires. To describe a circuit, we draw a **circuit diagram**, in which each element is replaced by a standard symbol (Figure 21.1). A circuit diagram shows how the elements are connected, not where they sit or what they look like: the lines are ideal wires with no resistance, and they may be stretched, bent, or redrawn in any way that keeps the same connections. Every point along a single wire, and everything attached directly to it, is at the same potential. A point where three or more wires meet is a **junction**, marked with a dot.

The battery and its emf. Charge flowing through a resistor always moves from high potential to low, losing potential energy. For a steady current to flow around a closed loop, something must lift the charge back up to high potential. A **battery** does this. Inside it, chemical reactions push positive charge from the negative terminal to the positive terminal, *against* the electric field there, doing work on it. (In a metal the carriers are electrons, pushed the other way, but the energy accounting is the same.) The battery is to charge what a pump is to water in a closed loop of pipe, or a ski lift to skiers: it raises them to a height from which they can run back down.

The work done by the source per unit charge that passes through it is called its **emf**, symbol $\mathcal{E}$:

Key Equation: Emf of a source

$$\mathcal{E} = \frac{W}{q}. \tag{21.1}$$

Its SI unit is the volt, one joule per coulomb. The name is short for “electromotive force,” a historical term that is misleading, since emf is not a force at all but an energy per unit charge; physicists now simply say “emf” (e-m-f). A battery with an emf of 1.5 V gives each coulomb that passes through it 1.5 J of energy. Batteries are just one example of a source of emf; the generators of Chapter 24, solar cells, and thermocouples are others.

An **ideal battery** maintains a potential difference between its terminals called its **terminal voltage**. This is exactly equal to its emf, however much current it delivers. Real batteries fall short of this, as Section 21.4 will show, but until then every battery is ideal: the potential of its positive terminal is higher than that of its negative terminal by exactly $\mathcal{E}$.

Example 21.1: The energy in a battery

An AA battery has an emf of $1.5\ \mathrm{V}$ and a capacity of about $2500\ \mathrm{mAh}$ (Chapter 20). How much energy can it deliver?

Solution

The capacity is the charge it can move through a circuit: $Q = (2.5\ \mathrm{A})(3600\ \mathrm{s}) = 9000\ \mathrm{C}$. Each coulomb gains $1.5\ \mathrm{J}$, so the energy is

$$W = Q\mathcal{E} = (9000\,\mathrm{C})(1.5\,\mathrm{V}) = \boxed{1.4 \times 10^4\,\mathrm{J}}. \tag{21.1.1}$$

That is enough to lift a $70\ \mathrm{kg}$ person about $20\ \mathrm{m}$, or to run a $1\ \mathrm{W}$ flashlight bulb for nearly four hours. By comparison, the same mass of gasoline, about $23\ \mathrm{g}$, holds nearly eighty times as much chemical energy, which is one reason electric cars need such large battery packs.

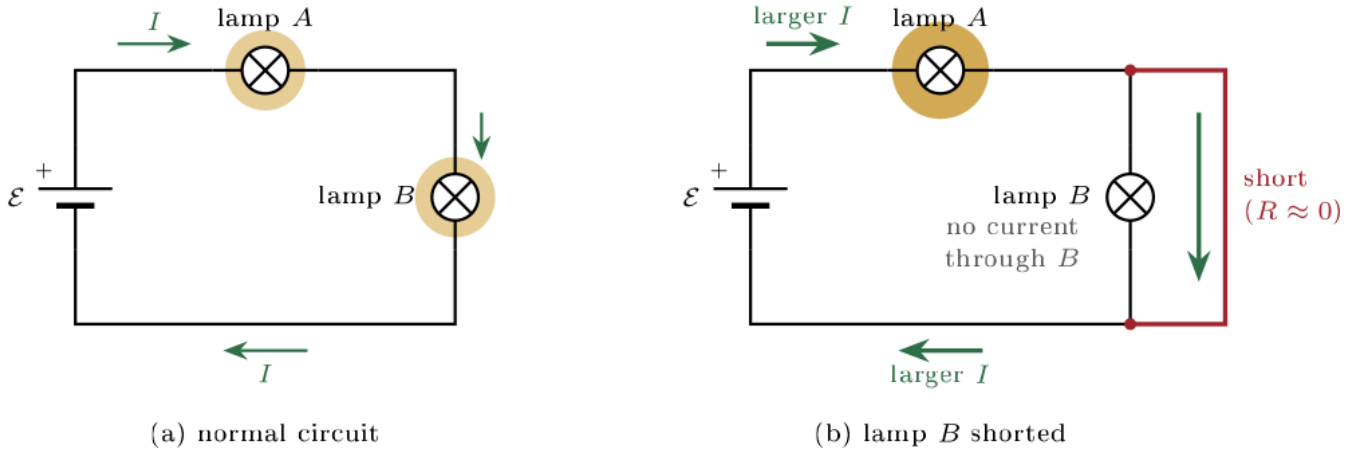


Figure 21.2. A short circuit. (a) Current flows around the whole loop, through the battery and both lamps. (b) A wire of negligible resistance connects the two ends of lamp B. The current takes this path instead, bypassing lamp B, which goes dark. With less total resistance in the circuit, the current through the battery and lamp A increases, and lamp A glows more brightly. A short directly across the battery would bypass everything, leaving only the battery's own small resistance to limit the current.

•Figure description

Two panels showing the same simple circuit: a battery on the left side of a rectangular loop, lamp A on the top side, and lamp B on the right side. In panel a, current I flows around the whole loop, through the battery and both lamps, and both lamps are lit. In panel b, a red wire of negligible resistance, labeled short, connects the two ends of lamp B, running around its outside. The current now flows through the short instead of through lamp B, so lamp B is dark and labeled no current. Lamp A is drawn brighter, and the current around the rest of the loop is labeled as larger.

Switches, open circuits, and short circuits. A **switch** opens or closes a gap in a wire. When the switch is open, there is no complete path, and no current can flow; the circuit is an **open circuit**. A **short circuit**, or **short**, results when a wire of negligible resistance connects two points of a circuit so that current bypasses the circuit elements between those points (Figure 21.2). A short across a battery draws an enormous current, limited only by the battery's own resistance, and can overheat the battery or the wire dangerously.

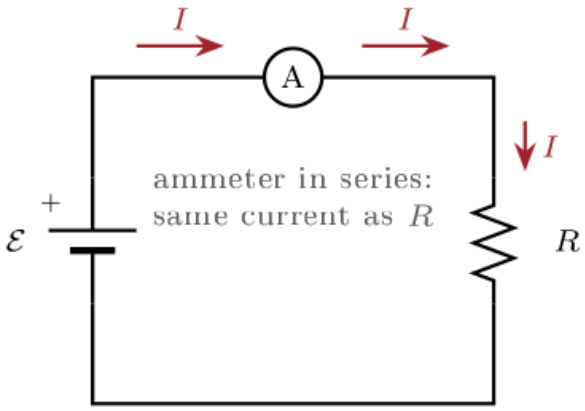


Figure 21.3. Measuring current. The circuit is broken and the ammeter inserted in series with the resistor, so that the current through the resistor must also pass through the ammeter. An ideal ammeter has zero resistance, so inserting it does not change the current.

•Figure description

A circuit diagram with a battery of emf script E on the left, positive terminal on top, connected in a rectangular loop to a resistor R on the right side. The top wire has been broken and an ammeter, a circle marked A, inserted in the gap, so that the ammeter and the resistor lie one after the other along the same single path. The current I, flowing clockwise, is shown by arrows on the wire entering the ammeter and on the wire leaving it, and continues down through the resistor, so the same current passes through the ammeter and the resistor.

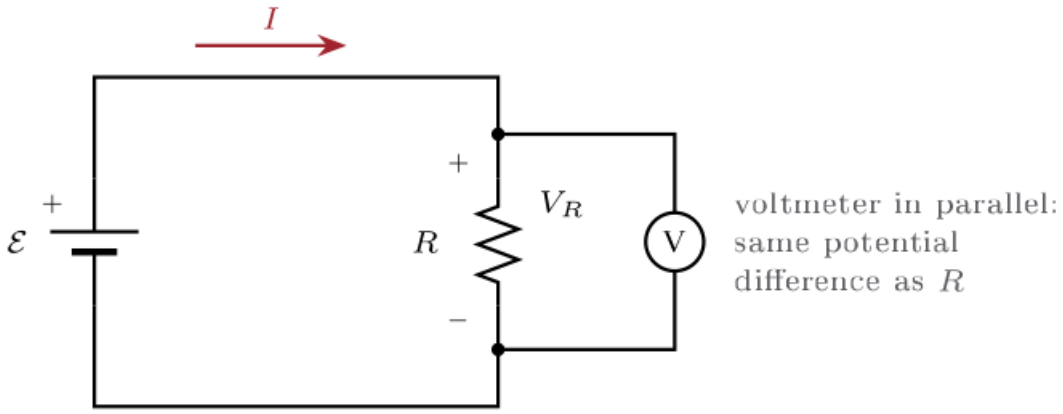


Figure 21.4. Measuring potential difference. The voltmeter is connected across the resistor, in parallel with it, without breaking the circuit; its two leads touch the two ends of the resistor, so it has the same potential difference V_R across it. An ideal voltmeter has infinite resistance, so it draws no current and does not disturb the circuit.

•Figure description

A circuit diagram with a battery of emf script E on the left, positive terminal on top, connected in a rectangular loop to a resistor R on the right side. No wire of the loop is broken. A voltmeter, a circle marked V, sits to the right of the resistor, and two extra wires connect its terminals to junction dots at the top and bottom ends of the resistor, so that the voltmeter and the resistor are side by side in parallel, spanning the same two points. The potential difference V sub R across the resistor is marked with a plus sign at its top end and a minus sign at its bottom end.

Measuring current and voltage. An **ammeter** measures the current through a wire. To do so, the current must flow through it, so it is connected *in series*: the wire is cut, and the ammeter is inserted in the gap (Figure 21.3). A **voltmeter** measures the potential difference between two points, so it is connected directly *across* them, in parallel with the element being measured, without cutting any wire (Figure 21.4). An ideal ammeter has zero resistance, so inserting it changes nothing; an ideal voltmeter has infinite resistance, so it draws no current. Real meters come close enough for most purposes; the digital multimeters found in every laboratory measure both, and resistance too.

Checkpoint 21.1

A student connects an ammeter directly across a resistor, the way a voltmeter should be connected. (a) Since the ammeter has almost no resistance, where does nearly all the current now flow? (b) Why might this damage the ammeter?

Answer

Answers to Checkpoints

Checkpoint 21.1. (a) Through the ammeter. Connected across the resistor, it is in parallel with it, and nearly all the current takes the almost-zero-resistance path, bypassing the resistor. (b) The ammeter has effectively short-circuited the resistor, so the current in the circuit is now limited only by whatever resistance remains elsewhere: perhaps just the battery's internal resistance. That current can be far larger than the meter is built to carry, blowing its fuse or burning out its coil.

This page titled 21.1: Circuit Elements and Diagrams is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.2: Kirchhoff's Laws and the Basic Circuit

Two rules, stated by Gustav Kirchhoff in 1845 while still a student, are enough to analyze any dc circuit. Both are conservation laws in disguise.

The junction rule. Charge does not pile up anywhere in a circuit carrying steady currents, so the total current flowing into any junction must equal the total current flowing out of it (Chapter 20). This is the **junction rule**:

Key Equation: Kirchhoff's junction rule

$$\sum I_{\text{in}} = \sum I_{\text{out}}. \tag{21.2}$$

The loop rule. The electric potential has a single value at each point (Chapter 18). So if you start at any point in a circuit, travel around any closed loop, adding up the changes in potential across each element you pass, and return to the starting point, you must arrive back at the same potential. The changes must add to zero. This is the **loop rule**:

Key Equation: Kirchhoff's loop rule

$$\sum_{\text{closed loop}} \Delta V = 0. \tag{21.3}$$

It is a statement of energy conservation: a charge carried around the loop gains as much potential energy in the sources of emf as it loses in the resistors.

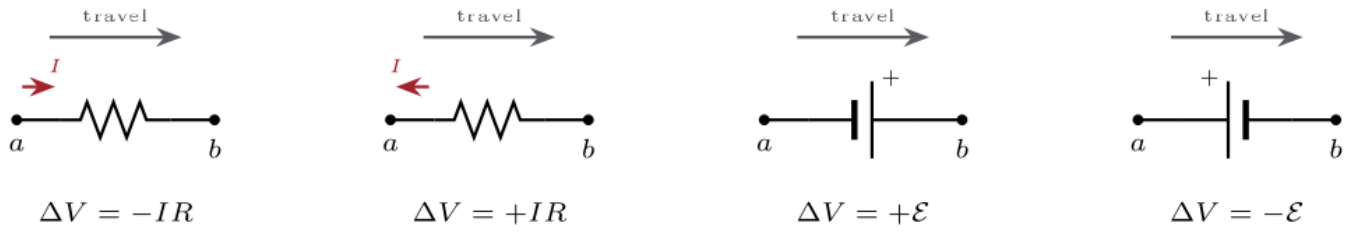


Figure 21.5. Sign rules for the loop rule, traveling from a to b. The potential drops by IR when you cross a resistor in the direction of the current, and rises by $\mathcal{E}$ when you cross a battery from its negative to its positive terminal.

•Figure description

Four panels, each showing one circuit element between a point a on the left and a point b on the right, with a gray arrow above showing that we travel from a to b. Panel 1: a resistor with current I flowing from a to b; the change in potential is minus I R. Panel 2: a resistor with current I flowing from b to a; the change is plus I R. Panel 3: a battery with its negative terminal toward a and positive terminal toward b; the change is plus script E. Panel 4: a battery with its positive terminal toward a and negative terminal toward b; the change is minus script E.

To use the loop rule, you need the sign of the potential change across each element, which depends on the direction in which you cross it (Figure 21.5):

- Crossing a resistor *in the direction of the current*, you go from high potential to low: $\Delta V = -IR$. Crossing it against the current, $\Delta V = +IR$.
- Crossing a battery from its negative terminal to its positive terminal, $\Delta V = +\mathcal{E}$. Crossing it from positive to negative, $\Delta V = -\mathcal{E}$. The direction of the current through the battery does not matter.

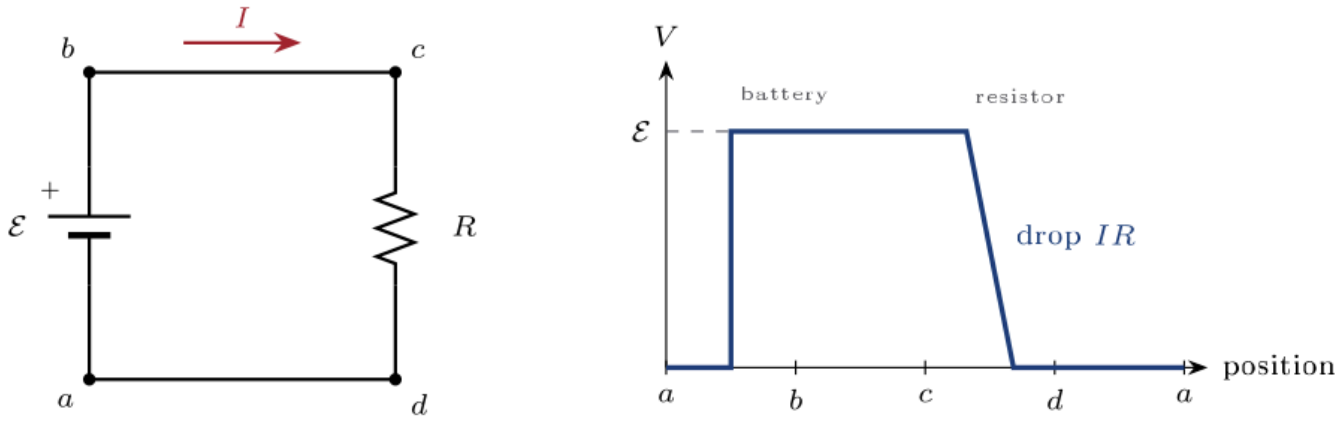


Figure 21.6. The basic circuit. Going clockwise from a, the potential rises by $\mathcal{E}$ across the battery and falls by IR across the resistor, returning to its starting value, so $\mathcal{E} = IR$.

•Figure description

Left: the basic circuit, a battery of emf script E on the left, positive terminal on top, and a resistor R on the right, joined by wires into a rectangle. The corners are labeled a at bottom left, b at top left, c at top right, and d at bottom right, and the current I flows clockwise. Right: a graph of the potential V against position around the loop, from a through b, c and d back to a. The potential is zero along the wire from a to the battery, jumps up by script E at the battery, stays constant at script E along the wire from b to c, falls steadily back to zero across the resistor, and stays at zero along the wire from d back to a.

The basic circuit. The simplest complete circuit is a single battery connected to a single resistor (Figure 21.6). Starting at point a and traveling clockwise, with the current, we rise by $\mathcal{E}$ crossing the battery and fall by IR crossing the resistor. The wires cause no change. The loop rule gives

$$\mathcal{E} - IR = 0, \qquad \text{so} \qquad I = \frac{\mathcal{E}}{R}. \tag{21.2.1}$$

The graph in Figure 21.6 plots the potential around the loop. It is like a roller coaster with a single lift hill: the battery raises the charge by $\mathcal{E}$, and all of that height is lost in the resistor. The whole potential difference of the battery appears across the resistor, because the ideal wires take none of it.

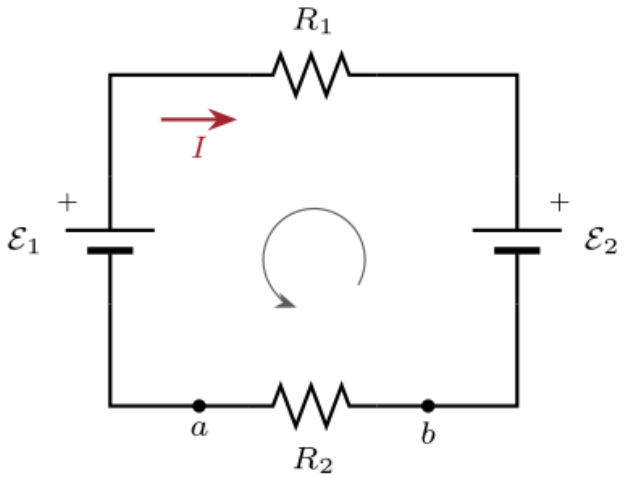


Figure 21.7. The circuit of Example 21.2. The two batteries oppose each other, and the stronger one wins: the current flows clockwise.

•Figure description

A single rectangular loop. On the left side is battery script E sub 1, positive terminal on top. On the right side is battery script E sub 2, also with its positive terminal on top, so the two batteries oppose each other. Resistor R sub 1 is in the top wire and resistor R sub 2 in the bottom wire. Point b is on the bottom wire to the right of R sub 2 and point a to its left. The current I is drawn clockwise, flowing to the right through R sub 1.

Example 21.2: A loop with two batteries

In the circuit of Figure 21.7, $\mathcal{E}_1 = 12\ \text{V}$, $\mathcal{E}_2 = 6.0\ \text{V}$, $R_1 = 4.0\ \Omega$, and $R_2 = 2.0\ \Omega$. Find the current, and the potential difference $V_b - V_a$ across R_2.

Solution

There is only one loop, so there is only one current I. Its direction is not obvious, since the batteries push in opposite directions, so *guess*: clockwise, as drawn. Starting at the bottom left corner and going clockwise, we cross $\mathcal{E}_1$ from $-$ to $+$ ($+\mathcal{E}_1$), R_1 with the current ($-IR_1$), $\mathcal{E}_2$ from $+$ to $-$ ($-\mathcal{E}_2$), and R_2 with the current ($-IR_2$):

$$\mathcal{E}_1 - IR_1 - \mathcal{E}_2 - IR_2 = 0 \quad \Longrightarrow \quad I = \frac{\mathcal{E}_1 - \mathcal{E}_2}{R_1 + R_2} = \frac{12 - 6.0}{4.0 + 2.0} = \boxed{1.0\,\text{A}}. \tag{21.2.2}$$

The answer is positive, so the guess was right: the current is clockwise. Had it come out negative, the current would simply flow the other way, with the same magnitude. Across R_2 the current flows from b to a, from high potential to low, so $V_b - V_a = +IR_2 = \boxed{2.0\,\text{V}}$.

Multiloop circuits. When a circuit has more than one loop, the currents in different branches differ. A *branch* is a stretch of circuit between two junctions, and it carries a single current. Label each branch current with a symbol and a guessed direction; write the junction rule at enough junctions (one fewer than the number of junctions); and write the loop rule for enough independent loops to have as many equations as unknown currents.

Note: A strategy for Kirchhoff's-law problems

1. Draw the circuit, and label every branch current with a symbol and an arrow. The direction is a guess; a negative answer means the current flows the other way.
2. Apply the junction rule to all but one of the junctions.
3. Choose loops, each containing at least one element not in the others, and apply the loop rule to each, using the sign rules of Figure 21.5.
4. Solve the equations. Check the answer by applying the loop rule to a loop you did not use.

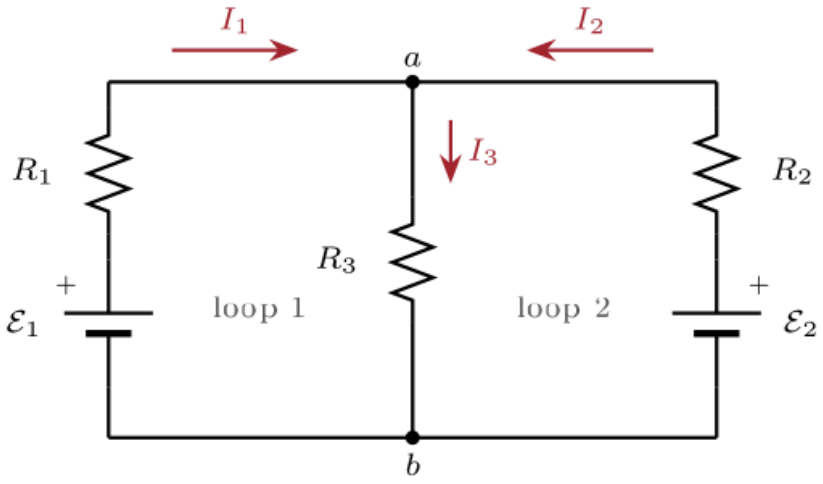


Figure 21.8. The two-loop circuit of Example 21.3. The current directions are guesses; the solution shows that I_2 actually flows the other way.

•Figure description

A two-loop circuit with three vertical branches connected between a top wire and a bottom wire. The left branch contains battery script E sub 1, positive terminal up, below resistor R sub 1. The middle branch contains resistor R sub 3. The right branch contains battery script E sub 2, positive terminal up, below resistor R sub 2. The middle branch meets the top wire at junction a and the bottom wire at junction b. Currents I sub 1 and I sub 2 flow along the top wire toward junction a from the left and right, and current I sub 3 flows down through R sub 3. The left loop is labeled loop 1 and the right loop, loop 2.

Example 21.3: A two-loop circuit

Find the three currents in the circuit of Figure 21.8, with $\mathcal{E}_1 = 12\ \text{V}$, $\mathcal{E}_2 = 6.0\ \text{V}$, $R_1 = 2.0\ \Omega$, $R_2 = 3.0\ \Omega$, and $R_3 = 6.0\ \Omega$.

Solution

The circuit has two junctions, a and b, and three branches. Take I_1 and I_2 upward through the outer branches, into a, and I_3 downward through R_3, out of a. The junction rule at a:

$$I_1 + I_2 = I_3. \tag{21.2.3}$$

Loop 1 (the left loop, clockwise from b): up through $\mathcal{E}_1$ and R_1, then down through R_3:

$$\mathcal{E}_1 - I_1R_1 - I_3R_3 = 0 \quad \Longrightarrow \quad 12 - 2.0I_1 - 6.0I_3 = 0. \tag{21.2.4}$$

Loop 2 (the right loop, counterclockwise from b): up through $\mathcal{E}_2$ and R_2, then down through R_3:

$$\mathcal{E}_2 - I_2R_2 - I_3R_3 = 0 \quad \Longrightarrow \quad 6.0 - 3.0I_2 - 6.0I_3 = 0. \tag{21.2.5}$$

From the loop equations, $I_1 = 6.0 - 3.0I_3$ and $I_2 = 2.0 - 2.0I_3$. Substituting into the junction rule, $8.0 - 5.0I_3 = I_3$, so

$$I_3 = \boxed{1.33\ \text{A}}, \qquad I_1 = \boxed{2.0\ \text{A}}, \qquad I_2 = \boxed{-0.67\ \text{A}}. \tag{21.2.6}$$

The negative sign means I_2 actually flows *downward* through $\mathcal{E}_2$, into its positive terminal. The stronger battery is driving current backward through the weaker one, charging it. Check with the outer loop, clockwise from b: $\mathcal{E}_1 - I_1R_1 + I_2R_2 - \mathcal{E}_2 = 12 - 4.0 - 2.0 - 6.0 = 0$. ✓

Checkpoint 21.2

In Figure 21.6, suppose you travel around the loop *counterclockwise*, against the current. Write the loop-rule equation you would get, and show that it gives the same current.

Answer

Energy and Power in a Circuit

The loop rule is an energy statement, and multiplying it by the current turns it into a statement about power. In the basic circuit, $\mathcal{E} = IR$; multiplied by I,

$$\mathcal{E}I = I^2R. \tag{21.2.7}$$

The left side is the rate at which the battery does work on the charge passing through it, and the right side is the rate at which the resistor turns electrical energy into heat (Section 20.4). The battery supplies exactly the power the resistor dissipates.

In general, a source of emf delivers power

Key Equation: Power delivered by a source of emf

$$P = \mathcal{E}I \tag{21.4}$$

when the current flows through it from its negative terminal to its positive terminal, the normal direction for a battery that is discharging. When a larger emf forces current through a battery the other way, into its positive terminal, the battery *absorbs* power $\mathcal{E}I$: it is being charged, and the electrical energy is stored as chemical energy. This is what happens when a car's alternator charges its battery, or when you plug in a phone. In every circuit, the power delivered by the sources equals the power absorbed by everything else.

Example 21.4: Energy bookkeeping

Account for the power in each element of the circuit of Example 21.2.

Solution

The current is $I = 1.0\ \text{A}$, clockwise. It flows through $\mathcal{E}_1$ from $-$ to $+$, so that battery delivers $\mathcal{E}_1 I = (12)(1.0) = 12\ \text{W}$. It flows through $\mathcal{E}_2$ from $+$ to $-$, so that battery absorbs $\mathcal{E}_2 I = 6.0\ \text{W}$: it is being charged. The resistors dissipate $I^2 R_1 = 4.0\ \text{W}$ and $I^2 R_2 = 2.0\ \text{W}$. The books balance:

$$\underbrace{12\,\text{W}}_{\text{delivered}} = \underbrace{6.0\,\text{W}}_{\text{stored in } \mathcal{E}_2} + \underbrace{4.0\,\text{W} + 2.0\,\text{W}}_{\text{heat}}. \tag{21.2.8}$$

Half of the power delivered by the first battery goes into charging the second, and half is lost as heat.

Answers to Checkpoints

Checkpoint 21.2. Start at a and go counterclockwise: along the bottom to the resistor, up through the resistor, along the top, and down through the battery. Crossing the resistor *against* the current, you go from low potential to high, a rise of $+IR$. Crossing the battery from its positive terminal to its negative terminal, you drop by $\mathcal{E}$. Back at a the changes must sum to zero: $+IR - \mathcal{E} = 0$, so $I = \mathcal{E}/R$, the same current. The loop rule does not care which way you travel, provided every sign follows the direction you actually cross each element.

This page titled 21.2: Kirchhoff's Laws and the Basic Circuit is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.3: Multiple Resistors in a Circuit: Series vs. Parallel

Kirchhoff's laws handle any circuit, but many circuits contain groups of resistors that can be replaced by a single **equivalent resistance**, one that draws the same current from the same potential difference. Two arrangements are basic. The rules for them are much like those for capacitors in Section 19.4, but reversed.

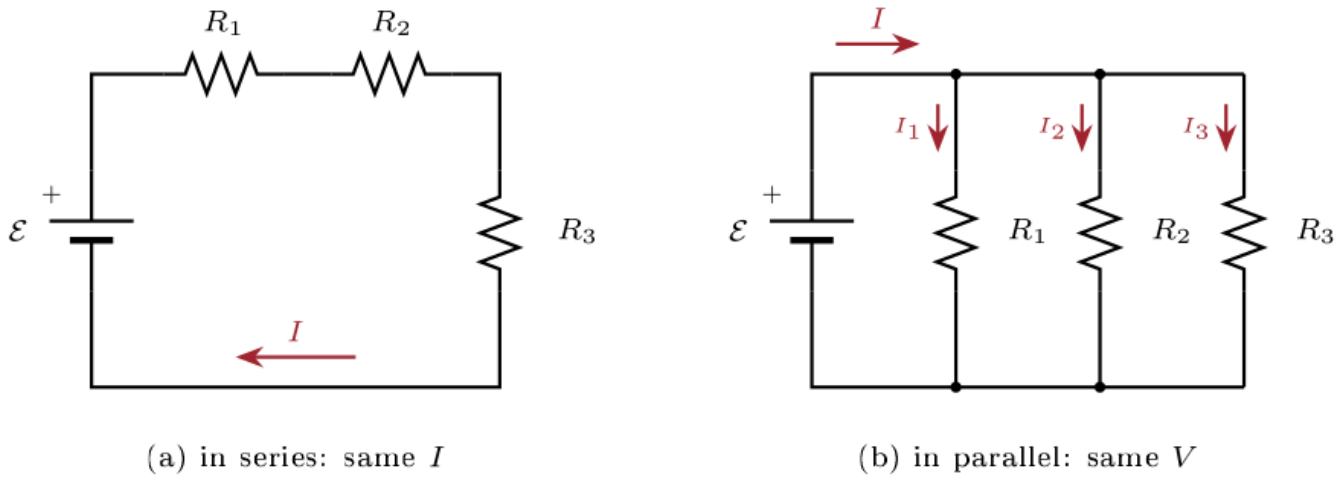


Figure 21.9. Resistors in combination. (a) In series, the same current passes through each resistor, and their potential differences add up to $\mathcal{E}$. (b) In parallel, each resistor has the same potential difference across it, and their currents add up to the battery current, $I = I_1 + I_2 + I_3$.

•Figure description

Two circuit diagrams. In panel a, a battery of emf script E is connected to three resistors R sub 1, R sub 2 and R sub 3 one after another around a single loop, R sub 1 and R sub 2 in the top wire and R sub 3 on the right side, so the same current I passes through each. In panel b, the same battery is connected to three resistors arranged side by side between a top wire and a bottom wire, each spanning the same pair of wires, so each has the same potential difference V across it; junction dots mark where each resistor joins the wires, and the battery current I divides into I sub 1, I sub 2 and I sub 3.

Resistors in series. Resistors connected one after another along a single path, with no junction between them, are in **series** (Figure 21.9a). There is only one path, so by the junction rule, *the same current* I passes through each. The loop rule says the potential differences across them add up to the battery's emf:

$$\mathcal{E} = IR_1 + IR_2 + IR_3 + \cdots = I(R_1 + R_2 + R_3 + \cdots). \tag{21.3.1}$$

A single resistor that draws the same current from the same emf has

Key Equation: Resistors in series

$$R_{\text{eq}} = R_1 + R_2 + R_3 + \cdots. \tag{21.5}$$

Resistances in series add, and the equivalent resistance is larger than any one of them. This is what we expect from $R = \rho L/A$: resistors in series are like a single longer resistor.

Resistors in parallel. Resistors connected side by side between the same two junctions are in **parallel** (Figure 21.9b). Each is connected directly across the same pair of wires, so each has *the same potential difference* V, and each carries its own current, $I_1 = V/R_1$, $I_2 = V/R_2$, and so on. By the junction rule, the total current from the battery is the sum:

$$I = I_1 + I_2 + I_3 + \cdots = V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots\right). \tag{21.3.2}$$

A single resistor that draws this total current at the same voltage has

Key Equation: Resistors in parallel

$$\frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots . \tag{21.6}$$

The equivalent resistance of resistors in parallel is always *smaller* than the smallest of them. Each new path in parallel gives the current another way through, like widening a pipe. For two resistors in parallel, equation (21.6) can be written $R_{\text{eq}} = R_1 R_2/(R_1 + R_2)$, "product over sum." Two equal resistors in parallel have half the resistance of either. The largest current flows through the smallest resistance, since $I = V/R$ for each.

Note: Household wiring is in parallel

The outlets and lights in a house are connected in parallel across the supply, so each device gets the full $120\ \text{V}$ (or $230\ \text{V}$ in much of the world), and each can be switched on or off without affecting the others. But every device added in parallel lowers the total resistance and raises the total current in the wires feeding the circuit. A circuit breaker or fuse, in series with the whole group, opens the circuit when that current exceeds a safe value, typically $15\ \text{A}$ or $20\ \text{A}$, before the wires in the walls overheat.

Example 21.5: Bulbs in series and in parallel

Two identical lamps, each with a resistance of $240\ \Omega$ when lit, are connected to a $120\ \text{V}$ supply. Find the power delivered to each lamp and the total power if they are connected (a) in parallel and (b) in series. Treat the resistance as constant.

Solution

(a) In parallel, each lamp has the full $120\ \text{V}$ across it, just as if it were alone, so each receives $P = V^2/R = (120)^2/240 = \boxed{60\ \text{W}}$, and the total is $120\ \text{W}$.

(b) In series, $R_{\text{eq}} = 480\ \Omega$, and the current is $I = 120/480 = 0.25\ \text{A}$. Each lamp receives $P = I^2R = (0.25)^2(240) = \boxed{15\ \text{W}}$, and the total is only $30\ \text{W}$. Each lamp gets half the voltage and half the current, and so a quarter of the power. (In practice the cooler filaments have lower resistance, so the lamps glow a little more than this, but still dimly.) Old strings of holiday lights were wired in series, which is why one burned-out bulb darkened the whole string.

Combinations. Many networks can be simplified step by step, replacing each group of resistors that is purely in series or purely in parallel by its equivalent, until only one resistor remains. Then the steps are retraced in reverse to find the current in, and voltage across, each original resistor.

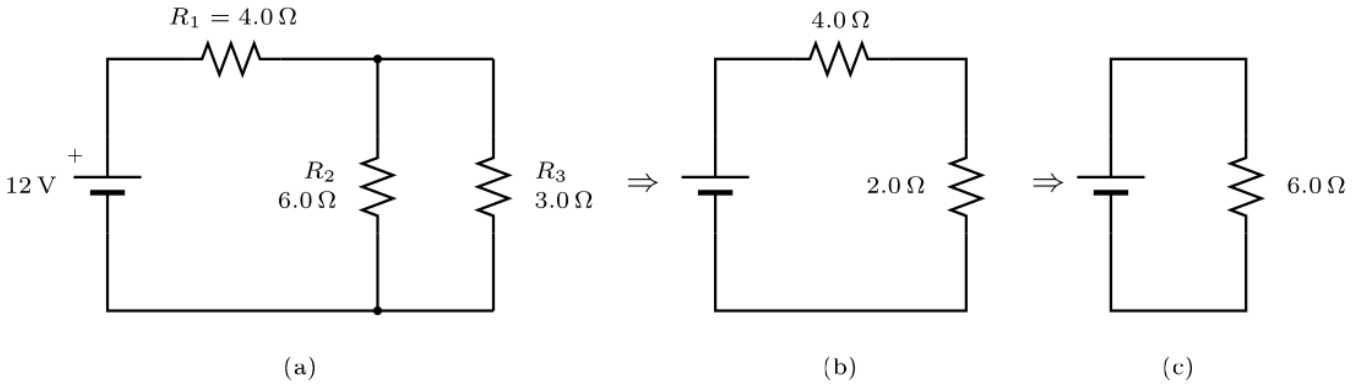


Figure 21.10. Simplifying the network of Example 21.6. (a) R_2 and R_3 are in parallel. (b) Their equivalent, $2.0\ \Omega$, is in series with R_1. (c) The whole network is equivalent to $6.0\ \Omega$.

•Figure description

Three circuit diagrams in a row, showing a network being simplified. Panel a: a 12 volt battery on the left, a 4.0 ohm resistor R sub 1 in the top wire, and two resistors in parallel on the right, R sub 2 of 6.0 ohms and R sub 3 of 3.0 ohms, each connected between the top and bottom wires. Panel b: the parallel pair has been replaced by a single 2.0 ohm resistor, in series with the 4.0 ohm resistor. Panel c: the whole network has been replaced by a single 6.0 ohm resistor across the battery.

Example 21.6: Simplifying a network

In the circuit of Figure 21.10a, $\mathcal{E} = 12\ \text{V}$, $R_1 = 4.0\ \Omega$, $R_2 = 6.0\ \Omega$, and $R_3 = 3.0\ \Omega$. Find the current in each resistor.

Solution

Step 1: Combine the parallel pair. R_2 and R_3 are connected between the same two junctions, so they are in parallel. By equation (21.6),

$$\frac{1}{R_{23}} = \frac{1}{R_2} + \frac{1}{R_3} = \frac{1}{6.0\,\Omega} + \frac{1}{3.0\,\Omega}. \tag{21.3.3}$$

To add the fractions, write them over the common denominator $6.0\ \Omega$:

$$\frac{1}{R_{23}} = \frac{1}{6.0\,\Omega} + \frac{2}{6.0\,\Omega} = \frac{3}{6.0\,\Omega} = \frac{1}{2.0\,\Omega}, \qquad \text{so} \qquad R_{23} = 2.0\,\Omega. \tag{21.3.4}$$

Remember to take the reciprocal at the end: $3/6.0$ is $1/R_{23}$, not R_{23}. The "product over sum" form gives the same result in one line, $R_{23} = R_2R_3/(R_2 + R_3) = (6.0)(3.0)/(6.0 + 3.0) = 18/9.0 = 2.0\ \Omega$. As expected, R_{23} is smaller than either R_2 or R_3.

Step 2: Combine in series. All of the battery current passes first through R_1 and then through the pair, so R_1 and R_{23} are in series (Figure 21.10b). By equation (21.5),

$$R_{\text{eq}} = R_1 + R_{23} = 4.0\,\Omega + 2.0\,\Omega = 6.0\,\Omega \tag{21.3.5}$$

(Figure 21.10c).

Step 3: Find the battery current. The single equivalent resistor has the whole emf across it, so

$$I = \frac{\mathcal{E}}{R_{\text{eq}}} = \frac{12\,\text{V}}{6.0\,\Omega} = 2.0\,\text{A}. \tag{21.3.6}$$

This current flows through R_1, since R_1 is in series with the battery: $I_1 = \boxed{2.0\,\text{A}}$.

Step 4: Work backward to the voltages. The same $2.0\ \text{A}$ passes through R_1 and through the equivalent resistor R_{23}, so

$$V_1 = IR_1 = (2.0\,\text{A})(4.0\,\Omega) = 8.0\,\text{V}, \qquad V_{23} = IR_{23} = (2.0\,\text{A})(2.0\,\Omega) = 4.0\,\text{V}. \tag{21.3.7}$$

Because R_2 and R_3 are in parallel, each has this same $4.0\ \text{V}$ across it. (Equivalently, by the loop rule, $V_{23} = \mathcal{E} - V_1 = 12 - 8.0 = 4.0\ \text{V}$.)

Step 5: Find the branch currents. Apply Ohm's law to each resistor of the pair separately, using the voltage across it:

$$I_2 = \frac{V_{23}}{R_2} = \frac{4.0\,\text{V}}{6.0\,\Omega} = \boxed{0.67\,\text{A}}, \qquad I_3 = \frac{V_{23}}{R_3} = \frac{4.0\,\text{V}}{3.0\,\Omega} = \boxed{1.33\,\text{A}}. \tag{21.3.8}$$

The smaller resistor carries the larger current: R_3 has half the resistance of R_2 and carries twice the current.

Check. At the junction where the current divides, the junction rule requires $I_2 + I_3 = I_1$: indeed $0.67 + 1.33 = 2.0\ \text{A}$. Around the loop through R_1 and either resistor of the pair, the loop rule requires the potential differences to add up to the emf: $8.0 + 4.0 = 12\ \text{V}$.

Example 21.7: A voltage divider

Two resistors in series across a source, $R_1 = 10\ \text{k}\Omega$ and $R_2 = 5.0\ \text{k}\Omega$ across a $9.0\ \text{V}$ battery, form a **voltage divider**. What is the potential difference across R_2?

Solution

The current is $I = \mathcal{E}/(R_1 + R_2)$, so the potential difference across R_2 is

$$V_2 = IR_2 = \mathcal{E}\,\frac{R_2}{R_1 + R_2} = (9.0)\frac{5.0}{15} = \boxed{3.0\,\text{V}}. \tag{21.3.9}$$

In series, the emf divides among the resistors in proportion to their resistances. Voltage dividers are everywhere in electronics, supplying a fraction of a voltage to part of a circuit. When one of the resistors is a variable resistor, such as the volume knob on an amplifier, turning the knob varies the output. When one is a sensor whose resistance depends on light or temperature, the output voltage becomes a measurement.

Checkpoint 21.3

Three identical bulbs are connected to a battery: bulb A in series with a parallel pair, B and C. (a) Rank the bulbs by brightness. (b) Bulb C is unscrewed. Does bulb A get brighter or dimmer? Does bulb B?

Answer

Answers to Checkpoints

Checkpoint 21.3. (a) All the current passes through A, and it splits equally between B and C, each carrying half. Since $P = I^2R$ with identical resistances, A is the brightest (four times as bright as either of the others) and B and C are equally bright. (b) With C removed, the circuit is just A and B in series, so the total resistance rises from $1.5R$ to $2R$ and the battery current *falls*: bulb A gets dimmer. Bulb B now carries that whole current, $\mathcal{E}/2R$, rather than half of the old current, $\frac{1}{2}\cdot\mathcal{E}/1.5R = \mathcal{E}/3R$, so B gets *brighter*.

This page titled 21.3: Multiple Resistors in a Circuit: Series vs. Parallel is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.4: Ideal vs. Real Batteries

Connect a voltmeter to a car battery, and it reads about $12.6\ \text{V}$. Turn the key, and while the starter motor cranks, the reading drops to perhaps $10\ \text{V}$, and the headlights dim. The battery's terminal voltage depends on the current it delivers. An ideal battery would not behave this way; a real one does, because the charge moving through the battery itself, through its electrolyte and electrodes, meets resistance.

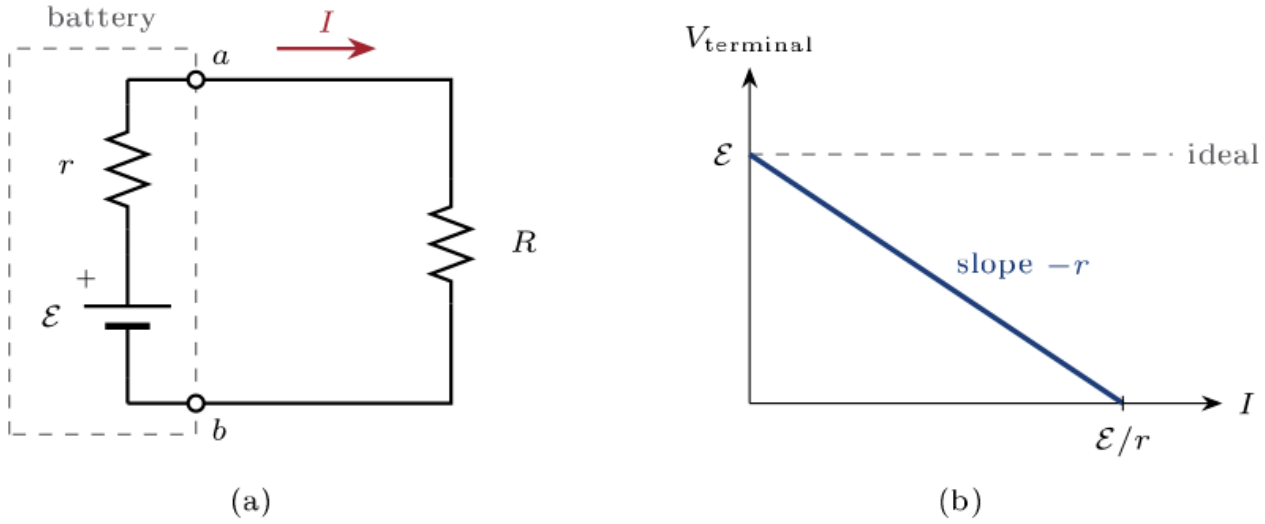


Figure 21.11. (a) A real battery modeled as an ideal emf $\mathcal{E}$ in series with an internal resistance r, connected to a load R. (b) Its terminal voltage falls linearly with the current it delivers, from $\mathcal{E}$ with no current to zero at the short-circuit current $\mathcal{E}/r$.

•Figure description

Panel a: a circuit diagram. A dashed box encloses an ideal battery of emf script E, positive terminal up, in series with a small resistor r, the internal resistance. The two terminals of the box are labeled a, at the top, and b, at the bottom, and are connected to an external load resistor R on the right. Current I flows clockwise. Panel b: a graph of terminal voltage against current. A horizontal dashed line at script E is labeled ideal battery. The real battery's terminal voltage is a straight line starting at script E at zero current and falling with slope minus r, reaching zero at the short-circuit current, script E over r.

A real battery is modeled as an ideal emf $\mathcal{E}$ in series with a small **internal resistance** r (Figure 21.11a). The two cannot be separated; they are inside the same case, and only the terminals a and b are accessible. When the battery delivers current I, going from its negative terminal to its positive, the potential rises by $\mathcal{E}$ across the emf and falls by Ir across the internal resistance, so the terminal voltage is

Key Equation: Terminal voltage of a real battery

$$V_{\text{terminal}} = \mathcal{E} - Ir. \tag{21.7}$$

Only when no current flows is the terminal voltage equal to the emf. This is how the emf is measured: with a high-resistance voltmeter, which draws almost no current. As the current rises, the terminal voltage falls along a straight line of slope $-r$ (Figure 21.11b), reaching zero for a short circuit, when the current is $I_{\text{sc}} = \mathcal{E}/r$. When a real battery is *charged*, current is forced into its positive terminal, and the terminal voltage is *larger* than the emf, $V_{\text{terminal}} = \mathcal{E} + Ir$.

Connected to an external load resistor R, the battery drives a current through both resistances in series:

$$I = \frac{\mathcal{E}}{R + r}. \tag{21.8}$$

If $R \gg r$, the internal resistance hardly matters, and the battery behaves as an ideal one. If R is comparable to r, much of the emf is lost inside the battery.

Example 21.8: Starting a car

A car battery has an emf of $12.6\ \text{V}$ and an internal resistance of $0.020\ \Omega$. (a) What is its terminal voltage while it supplies $150\ \text{A}$ to the starter motor? (b) At the same time, the headlights, which draw $10\ \text{A}$ at $12.6\ \text{V}$, are on. By what fraction does their power drop?

Solution

(a) $V_{\text{terminal}} = \mathcal{E} - Ir = 12.6 - (150)(0.020) = \boxed{9.6\,\text{V}}$. (The headlights' own $10\ \text{A}$ lowers it slightly more; we neglect that.)

(b) The headlights have resistance $R = 12.6/10 = 1.26\ \Omega$, and at constant resistance their power V^2/R is proportional to V^2. The power falls to $(9.6/12.6)^2 = 0.58$ of its normal value, a drop of about $\boxed{42\%}$. The headlights visibly dim while the engine cranks. On a cold morning, when the chemical reactions are slower, the internal resistance rises and the effect is worse, which is why cars are hardest to start in winter.

Getting the most power from a battery. The power delivered to the load is

$$P = I^2R = \frac{\mathcal{E}^2 R}{(R+r)^2}. \tag{21.4.1}$$

It is zero for $R = 0$ (no voltage across the load) and falls toward zero for very large R (almost no current). In between it has a maximum. Setting $dP/dR = 0$,

$$\frac{dP}{dR} = \mathcal{E}^2\,\frac{(R+r)^2 - 2R(R+r)}{(R+r)^4} = \mathcal{E}^2\,\frac{r-R}{(R+r)^3} = 0, \tag{21.4.2}$$

so the power to the load is greatest when the load resistance equals the internal resistance, $R = r$. This result, called **impedance matching** in its more general form, is used to connect amplifiers to loudspeakers and antennas to receivers, where getting the most power to the load is the aim. It is not how a power company runs its grid: at $R = r$, half of the power is wasted inside the source. Where efficiency matters, the load resistance is made much larger than the source's.

Example 21.9: Matching a load

A $9.0\ \text{V}$ battery has an internal resistance of $2.0\ \Omega$. (a) What is the most power it can deliver to an external resistor, and what resistor does this require? (b) What power does it deliver to a $20\ \Omega$ resistor, and what fraction of the battery's output is that?

Solution

(a) The maximum is at $R = r = 2.0\ \Omega$. Then $I = \mathcal{E}/(2r) = 2.25\ \text{A}$, and $P_{\text{max}} = I^2 r = \mathcal{E}^2/(4r) = (9.0)^2/8.0 = \boxed{10\,\text{W}}$.

(b) With $R = 20\ \Omega$, $I = 9.0/22 = 0.41\ \text{A}$, and $P = I^2R = \boxed{3.3\,\text{W}}$. The battery's total output is $\mathcal{E}I = 3.7\ \text{W}$, so the load receives $R/(R+r) = 20/22 = \boxed{91\%}$ of it. Less power, but far less waste.

Checkpoint 21.4

A flashlight's batteries are getting old, and their internal resistance has grown. (a) A voltmeter connected across them, with the flashlight off, still reads nearly their full emf. Why? (b) Why is the bulb dim when the flashlight is switched on?

Answer

Answers to Checkpoints

Checkpoint 21.4. (a) A good voltmeter draws almost no current, so there is almost no voltage drop Ir across the internal resistance, and the terminal voltage $\mathcal{E} - Ir$ is nearly the full emf. (b) With the bulb connected, a substantial current flows, and the large internal resistance now causes a large drop Ir inside the battery. The terminal voltage delivered to the bulb falls well below the emf, and the bulb receives much less power.

This page titled 21.4: Ideal vs. Real Batteries is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.5: RC Circuits

So far every current has been steady. A circuit containing a resistor and a capacitor, an **RC circuit**, is different. When a capacitor is connected to a battery, charge flows onto its plates, but as the charge builds up, so does the capacitor's own potential difference, $v = q/C$ (Chapter 19), and it opposes the battery. The current starts large and dies away. We use lowercase q, v, and i for quantities that change with time.

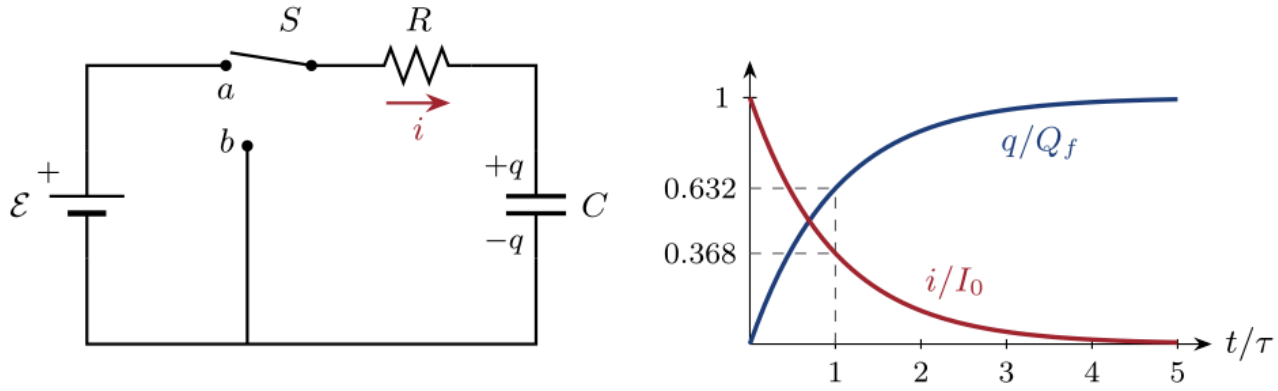


Figure 21.12. Left: an RC circuit. With the switch at a the capacitor charges; at b it discharges through R. Right: while charging, the capacitor's charge rises toward its final value $Q_f = C\mathcal{E}$, and the current falls from $I_0 = \mathcal{E}/R$, both on the time scale $\tau = RC$.

•Figure description

Left: an RC circuit. A battery of emf script E, positive terminal up, is on the left side. The top wire from the battery runs to contact a of a two-position switch S. The switch's pivot connects through resistor R, in the top wire, to capacitor C on the right side, whose lower lead returns along the bottom wire to the battery. A second switch contact, b, is connected straight down to the bottom wire. In position a the battery, resistor and capacitor form one loop; in position b the battery is cut out and the resistor and capacitor form a loop by themselves. Right: a graph against time in units of the time constant, from 0 to 5. The charge on the capacitor while charging, as a fraction of its final value, rises from zero along a curve that bends over toward 1; at one time constant it has reached 0.632, marked with dashed lines. The current, as a fraction of its initial value, falls from 1 toward zero along a decaying curve; at one time constant it is 0.368.

The circuit of Figure 21.12 can do both jobs. With the switch in position a, the battery charges the capacitor through the resistor. With the switch in position b, the battery is removed from the circuit, and the charged capacitor drives current back through the resistor and discharges.

Charging a Capacitor

Start with the capacitor uncharged and throw the switch to a at time $t = 0$. At a later time t, the capacitor carries charge q and the current is i. Going clockwise around the loop, the potential rises by $\mathcal{E}$ across the battery, falls by iR across the resistor, and falls by q/C across the capacitor, from its positive plate to its negative one. The loop rule gives

$$\mathcal{E} - iR - \frac{q}{C} = 0. \tag{21.9}$$

At the instant the switch closes, $q = 0$, and the capacitor has no potential difference: it acts like a plain wire, and the current is $I_0 = \mathcal{E}/R$, as if the capacitor were not there. Long afterward, the capacitor is full and the current stops, $i = 0$, so its potential difference equals the emf and its charge is $Q_f = C\mathcal{E}$: it acts like a break in the wire. Between these limits, the current is the rate at which charge arrives on the capacitor, $i = dq/dt$, and equation (21.9) becomes a differential equation for q:

$$\frac{dq}{dt} = \frac{\mathcal{E}}{R} - \frac{q}{RC} = -\frac{q - C\mathcal{E}}{RC}. \tag{21.5.1}$$

Separating the variables and integrating from $q = 0$ at $t = 0$,

$$\int_0^q \frac{dq'}{q' - C\mathcal{E}} = -\frac{1}{RC}\int_0^t dt' \quad \Longrightarrow \quad \ln\!\left(\frac{C\mathcal{E} - q}{C\mathcal{E}}\right) = -\frac{t}{RC}. \tag{21.5.2}$$

Taking the exponential of both sides and solving for q, then differentiating to find i:

Key Equation: Charging a capacitor

$$q(t) = C\mathcal{E}\left(1 - e^{-t/RC}\right), \qquad i(t) = \frac{\mathcal{E}}{R}\, e^{-t/RC}. \tag{21.10}$$

The charge rises toward $Q_f = C\mathcal{E}$, and the current decays from $I_0 = \mathcal{E}/R$ toward zero (Figure 21.12). Both change on the time scale set by the product

Key Equation: Time constant

$$\tau = RC, \tag{21.11}$$

the **time constant** of the circuit. (Check the units: $\Omega \cdot \mathrm{F} = (\mathrm{V/A})(\mathrm{C/V}) = \mathrm{C/A} = \mathrm{s}$.) After one time constant the charge has reached $1 - e^{-1} = 63.2\%$ of its final value and the current has fallen to $e^{-1} = 36.8\%$ of its initial value. After 2τ the charge is at 86%, after 3τ at 95%, and after 5τ at over 99%; in practice, a capacitor is "fully" charged after a few time constants. A large resistance slows the charging because it limits the current; a large capacitance slows it because more charge is needed to fill it.

The same exponential approach to a limit appeared in Section 6.2, for an object falling against linear drag: $v(t) = v_t(1 - e^{-bt/m})$. The mathematics is identical because the physics has the same shape. In each case the driving influence (gravity, the battery) is opposed by something that grows as the system approaches its final state (drag, the capacitor's voltage), and the rate of approach is proportional to how far there is still to go.

Where the energy goes. While the capacitor charges, the battery moves a total charge $Q_f = C\mathcal{E}$ through the emf, doing work $Q_f\mathcal{E} = C\mathcal{E}^2$. The capacitor ends up storing $U = \frac{1}{2}C\mathcal{E}^2$ (equation (19.5)), exactly half. The other half is dissipated in the resistor, as we can check directly:

$$\int_0^\infty i^2 R\, dt = \frac{\mathcal{E}^2}{R}\int_0^\infty e^{-2t/RC}\, dt = \frac{\mathcal{E}^2}{R}\cdot\frac{RC}{2} = \frac{1}{2}C\mathcal{E}^2. \tag{21.5.3}$$

Remarkably, the result does not depend on R. A smaller resistor allows a larger current for a shorter time, and the heat produced comes out the same.

Example 21.10: Charging a capacitor

A $1.5\ \mu\mathrm{F}$ capacitor is charged through a $2.0\ \mathrm{M\Omega}$ resistor by a $9.0\ \mathrm{V}$ battery. Find (a) the time constant, (b) the initial current, (c) the charge on the capacitor $5.0\ \mathrm{s}$ after the switch is closed, and (d) the time for the capacitor to reach 99% of its final charge.

Solution

(a) $\tau = RC = (2.0 \times 10^6)(1.5 \times 10^{-6}) = \boxed{3.0\,\mathrm{s}}$.

(b) $I_0 = \mathcal{E}/R = 9.0/(2.0 \times 10^6) = \boxed{4.5\,\mu\mathrm{A}}$.

(c) The final charge is $Q_f = C\mathcal{E} = 13.5\ \mu\mathrm{C}$, so

$$q = Q_f\left(1 - e^{-5.0/3.0}\right) = (13.5)(1 - 0.189) = \boxed{11\,\mu\mathrm{C}}. \tag{21.5.4}$$

(d) Setting $1 - e^{-t/\tau} = 0.99$ gives $e^{-t/\tau} = 0.01$, so $t = \tau \ln 100 = (3.0)(4.61) = \boxed{14\,\mathrm{s}}$, a little under five time constants.

Discharging a Capacitor

Now let the capacitor start with charge Q_0, and move the switch in Figure 21.12 to position b at $t = 0$. The battery is out of the circuit, and the capacitor alone drives current around the loop through R, from its positive plate through the resistor to its negative plate. Going around the loop in the direction of this current, the potential falls by iR in the resistor and rises by q/C across the capacitor, so

$$\frac{q}{C} - iR = 0. \tag{21.5.5}$$

Now the current *drains* the capacitor, so $i = -dq/dt$, and

$$\frac{dq}{dt} = -\frac{q}{RC} \quad \Longrightarrow \quad \int_{Q_0}^{q} \frac{dq'}{q'} = -\frac{1}{RC}\int_0^t dt' \quad \Longrightarrow \quad \ln\frac{q}{Q_0} = -\frac{t}{RC}. \tag{21.5.6}$$

Both the charge and the current decay exponentially, with the same time constant $\tau = RC$:

Key Equation: Discharging a capacitor

$$q(t) = Q_0\, e^{-t/RC}, \qquad i(t) = \frac{Q_0}{RC}\, e^{-t/RC} = \frac{V_0}{R}\, e^{-t/RC}, \tag{21.12}$$

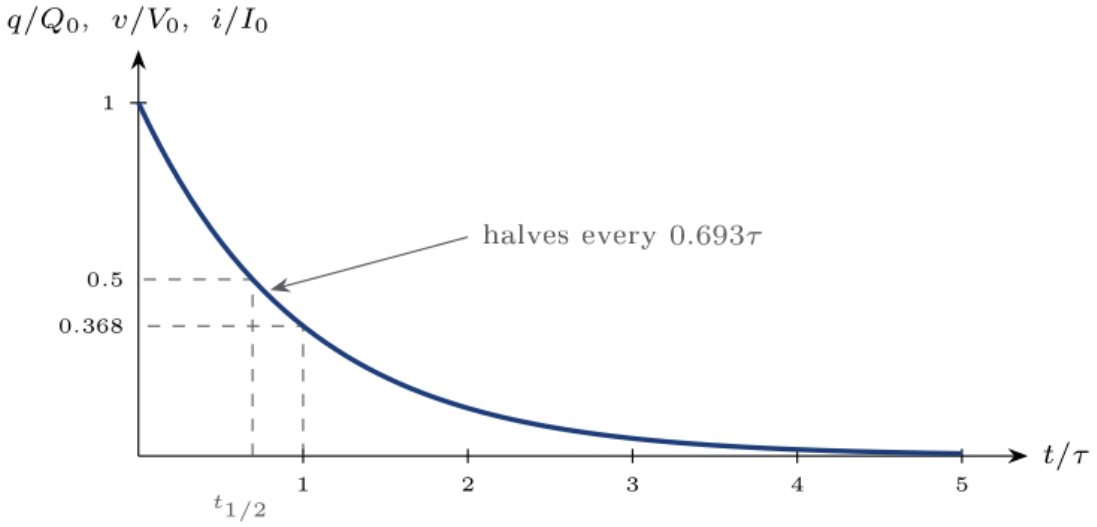


Figure 21.13. A capacitor discharging through a resistor. Its charge, its voltage, and the current all decay as $e^{-t/RC}$, falling to 36.8% after one time constant and to half after $t_{1/2} = 0.693\ RC$.

•Figure description

A graph of the charge on a discharging capacitor, as a fraction of its initial charge, against time in units of the time constant, from 0 to 5. The curve starts at 1 and decays exponentially toward zero. Dashed lines mark that the charge has fallen to one half at t over tau equal to 0.693, the half-life, and to 0.368 at t over tau equal to 1. The same curve describes the capacitor's voltage and the current.

where $V_0 = Q_0/C$ is the capacitor's initial potential difference. The capacitor's voltage $v = q/C$ decays in the same way (Figure 21.13). After one time constant, 36.8% of the charge remains. The time for half of it to drain away, the **half-life** of the discharge, follows from $e^{-t/\tau} = \frac{1}{2}$:

$$t_{1/2} = \tau \ln 2 = 0.693\, RC. \tag{21.13}$$

In each half-life the charge halves again, whatever it was to begin with. This is the same law that governs radioactive decay, where the number of undecayed nuclei halves in each half-life (Chapter 43). The energy stored in the capacitor, $q^2/2C$, decays twice as fast, as $e^{-2t/RC}$, and all of it ends up as heat in the resistor.

Example 21.11: A defibrillator

A defibrillator stores energy in a $32\ \mu\text{F}$ capacitor charged to $5000\ \text{V}$ and discharges it through the patient's chest, whose resistance is about $50\ \Omega$. Find (a) the stored energy, (b) the time constant and the initial current, and (c) the energy remaining in the capacitor after $2.0\ \text{ms}$.

Solution

(a) $U_0 = \frac{1}{2}CV_0^2 = \frac{1}{2}(32\times10^{-6})(5000)^2 = \boxed{400\,\text{J}}$.

(b) $\tau = RC = (50)(32\times10^{-6}) = \boxed{1.6\,\text{ms}}$, and $I_0 = V_0/R = 5000/50 = \boxed{100\,\text{A}}$, a huge current for a few milliseconds.

(c) The energy decays as $e^{-2t/\tau}$:

$$U = U_0\,e^{-2t/\tau} = (400)\,e^{-2(2.0)/1.6} = (400)\,e^{-2.5} = \boxed{33\,\text{J}}. \tag{21.5.7}$$

More than nine-tenths of the energy is delivered within $2\ \text{ms}$. The short, intense pulse depolarizes the heart muscle all at once, giving its natural pacemaker the chance to restart a normal rhythm. (Some defibrillators shape the pulse with an inductor as well, Chapter 25, but the RC discharge is the heart of it.)

RC circuits at work. Because $\tau = RC$ can be set anywhere from nanoseconds to minutes by choosing R and C, RC circuits are the standard way to make a circuit keep time. A capacitor charging toward a threshold voltage triggers the flash of a blinking warning light, the sweep of an intermittent windshield wiper, or the pulse of a cardiac pacemaker; then it discharges and the cycle repeats. The same circuits smooth out fluctuating voltages in power supplies, filter noise from audio signals, and, with τ of a few picoseconds, limit how fast the wires on a computer chip can carry a signal.

Checkpoint 21.5

A capacitor is charging through a resistor. (a) What fraction of its final charge does it hold after two time constants? (b) What fraction of its final *energy*? (c) If the resistance is doubled, how do the time constant, the initial current, and the final charge change?

Answer

Answers to Checkpoints

Checkpoint 21.5. (a) $q = Q_f\left(1 - e^{-t/\tau}\right)$, so at $t = 2\tau$ the fraction is $1 - e^{-2} = 1 - 0.135 = 0.865$, about 86%. (b) Stored energy is $U = q^2/2C$, proportional to q^2, so the fraction of the final energy is $(0.865)^2 \approx 0.75$: about 75%. (c) The time constant $\tau = RC$ doubles; the initial current $\mathcal{E}/R$ halves; and the final charge $C\mathcal{E}$ is unchanged, since the resistor only slows the charging, not its end point.

This page titled 21.5: RC Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.S: DC Circuits (Summary)

Summary

- A **circuit diagram** shows how elements are connected, using standard symbols; each wire is at a single potential. A source's **emf** $\mathcal{E}$ is the work it does per unit charge, in volts. Ammeters go in series; voltmeters go across.
- **Kirchhoff's rules**: the **junction rule**, $\sum I_{\text{in}} = \sum I_{\text{out}}$ (charge conservation), and the **loop rule**, $\sum \Delta V = 0$ around any closed loop (energy conservation). Crossing a resistor with the current, $\Delta V = -IR$; crossing a battery from $-$ to $+$, $\Delta V = +\mathcal{E}$.
- A source delivers power $\mathcal{E}I$ when current leaves its positive terminal, and absorbs it when being charged. The power supplied equals the power absorbed.
- Resistors in **series** carry the same current: $R_{\text{eq}} = R_1 + R_2 + \cdots$. Resistors in **parallel** have the same voltage: $1/R_{\text{eq}} = 1/R_1 + 1/R_2 + \cdots$. A series pair divides the voltage in proportion to the resistances.
- A real battery is an emf with an **internal resistance** r: its terminal voltage is $\mathcal{E} - Ir$, and it drives $I = \mathcal{E}/(R + r)$ through a load. The load receives the most power when $R = r$.
- In an **RC circuit**, charge and current change exponentially with **time constant** $\tau = RC$. Charging: $q = C\mathcal{E}(1 - e^{-t/RC})$, $i = (\mathcal{E}/R)e^{-t/RC}$; half the battery's work is stored and half becomes heat. Discharging: $q = Q_0 e^{-t/RC}$, with half-life $0.693\ RC$.

Key Equations

Emf

$$\mathcal{E} = \frac{W}{q} \tag{21.S.1}$$

Kirchhoff's rules

$$\sum I_{\text{in}} = \sum I_{\text{out}}, \qquad \sum_{\text{loop}} \Delta V = 0 \tag{21.S.2}$$

Power from a source of emf

$$P = \mathcal{E}I \tag{21.S.3}$$

Resistors in series and in parallel

$$R_{\text{eq}} = R_1 + R_2 + \cdots, \qquad \frac{1}{R_{\text{eq}}} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots \tag{21.S.4}$$

Real battery

$$V_{\text{terminal}} = \mathcal{E} - Ir, \qquad I = \frac{\mathcal{E}}{R + r} \tag{21.S.5}$$

Charging a capacitor

$$q = C\mathcal{E}\left(1 - e^{-t/RC}\right), \qquad i = \frac{\mathcal{E}}{R}\, e^{-t/RC} \tag{21.S.6}$$

Discharging a capacitor

$$q = Q_0\, e^{-t/RC}, \qquad i = \frac{V_0}{R}\, e^{-t/RC} \tag{21.S.7}$$

Time constant and half-life

$$\tau = RC, \qquad t_{1/2} = RC \ln 2 \tag{21.S.8}$$

This page titled 21.S: DC Circuits (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

21.E: DC Circuits (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Batteries are ideal unless an internal resistance is given.

Section 21.1 — Circuit Elements and Diagrams

21.1 ★ How much work does a $9.0\ \text{V}$ battery do in moving $0.25\ \text{C}$ of charge from its negative terminal to its positive terminal?

21.2 ★ *[Real-world: phones.]* A phone battery has an emf of $3.7\ \text{V}$ and a capacity of $4000\ \text{mAh}$. How much energy does it store, in joules and in watt-hours?

21.3 ★★ *[Conceptual.]* Draw a circuit diagram for a flashlight with two batteries in series, a switch, and a bulb. Show where you would connect an ammeter to measure the bulb's current, and a voltmeter to measure the voltage across the bulb.

Section 21.2 — Kirchhoff's Laws and the Basic Circuit

21.4 ★ A $6.0\ \text{V}$ battery is connected in a single loop with a $10\ \Omega$ and a $20\ \Omega$ resistor. Use the loop rule to find the current and the potential difference across each resistor.

21.5 ★★ A single loop contains a $9.0\ \text{V}$ battery, a $1.5\ \text{V}$ battery connected so as to oppose it, and resistors of $5.0\ \Omega$ and $10\ \Omega$. Find the current. What would it be if the $1.5\ \text{V}$ battery were turned around?

21.6 ★★ Repeat Example 21.3 with $\mathcal{E}_1 = 10\ \text{V}$, $\mathcal{E}_2 = 5.0\ \text{V}$, $R_1 = R_2 = 5.0\ \Omega$, and $R_3 = 10\ \Omega$. Which battery, if either, is being charged? Find the potential of junction a relative to junction b.

21.7 ★★ *[Conceptual.]* In Example 21.3, three loops can be drawn, but only two loop equations were used. Show that the loop equation for the outer loop is the difference of the other two, and so contains no new information.

21.8 ★★★ Explain why the loop rule would fail if the electric field in a circuit were not conservative, that is, if the work done by the field on a charge carried around a closed path were not zero. (Chapter 24 shows that a changing magnetic field produces exactly such a field, and the loop rule must then be modified.)

Section 21.2.1 — Energy and Power in a Circuit

21.9 ★★ In Example 21.2, the $6.0\ \text{V}$ battery is turned around so that the two batteries aid each other. Find the current and the power delivered or absorbed by each element, and check that the books balance.

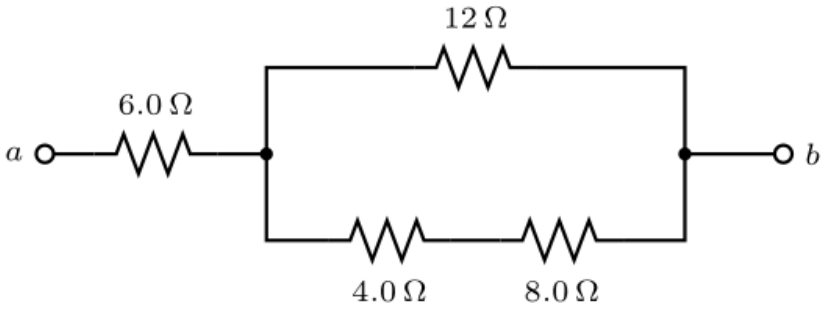
21.10 ★★ *[Real-world: battery charging.]* A charger forces a current of $5.0\ \text{A}$ into the positive terminal of a car battery whose emf is $12.0\ \text{V}$ and internal resistance $0.10\ \Omega$. (a) At what rate is energy stored in the battery? (b) At what rate is heat produced inside it? (c) What is the terminal voltage?

Section 21.3 — Multiple Resistors in a Circuit: Series vs. Parallel

21.11 ★ Find the equivalent resistance of $10\ \Omega$, $20\ \Omega$, and $30\ \Omega$ resistors (a) in series and (b) in parallel.

21.12 ★ How many $100\ \Omega$ resistors must be connected in parallel to give an equivalent resistance of $12.5\ \Omega$?

21.13 ★★ Find the equivalent resistance between terminals a and b of the network below. If $24\ \text{V}$ is applied between a and b, find the current in the $12\ \Omega$ resistor.



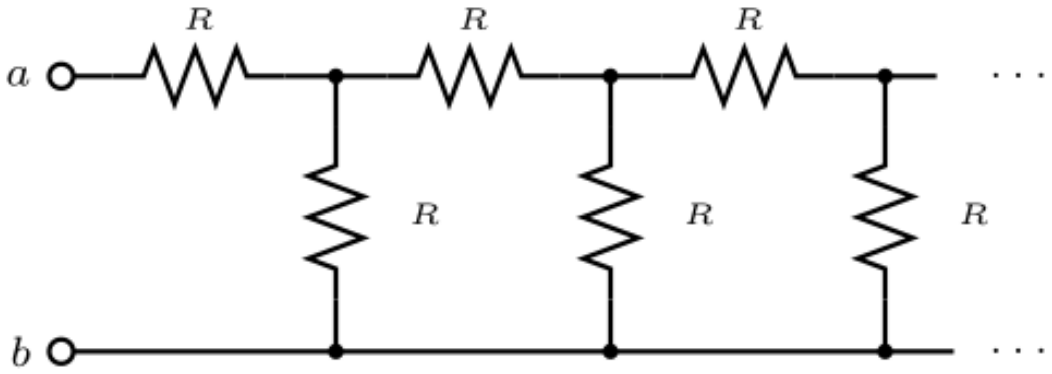
•Figure description
A resistor network between an open terminal a on the left and an open terminal b on the right. From a, a 6.0 ohm resistor leads to a junction. From that junction two branches run to a second junction: the upper branch contains a single 12 ohm resistor, and the lower branch contains a 4.0 ohm and an 8.0 ohm resistor in series. From the second junction a wire leads to terminal b.

21.14 ★★ *[Conceptual.]* For the three bulbs of the checkpoint in Section 21.3 (A in series with the parallel pair B and C), with a $12\ \text{V}$ battery and bulbs of $6.0\ \Omega$ each, find the power in each bulb before and after C is removed. Do your numbers agree with your answers to the checkpoint?

21.15 ★★ *[Real-world: measurement.]* A voltage divider consists of two $10\ \text{k}\Omega$ resistors across a $9.0\ \text{V}$ battery. (a) What is the voltage across each? (b) A voltmeter with a resistance of $10\ \text{k}\Omega$ is connected across one of the resistors. What does it read? (c) What would a voltmeter of resistance $10\ \text{M}\Omega$ read? Explain why a good voltmeter must have a very high resistance.

21.16 ★★ *[Real-world: household circuits.]* A $1500\ \text{W}$ space heater, a $1200\ \text{W}$ hair dryer, and a $100\ \text{W}$ lamp are all plugged into outlets on the same $120\ \text{V}$ circuit, which has a $15\ \text{A}$ breaker. Find the total current. Does the breaker trip? What is the equivalent resistance of the three appliances?

21.17 ★★★ Find the equivalent resistance between a and b of the infinite ladder of identical resistors R shown. *(Hint: adding one more section to the front of an infinite ladder leaves it unchanged, so $R_{\text{eq}} = R + (R \parallel R_{\text{eq}})$.)*

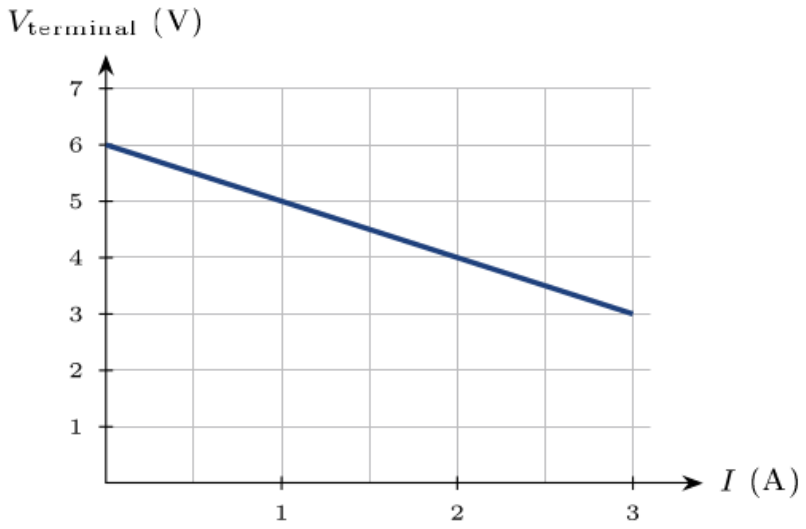


•Figure description
An infinite ladder network of identical resistors R. Two open terminals, a at top left and b at bottom left, begin a top wire and a bottom wire. Along the top wire, resistors R are repeated in series; after each one, a rung resistor R connects the top wire down to the bottom wire. Three sections are drawn, followed by dots indicating that the pattern continues forever.

Section 21.4 — Ideal vs. Real Batteries

21.18 ★ A battery with an emf of $1.50\ \text{V}$ and an internal resistance of $0.30\ \Omega$ is connected to a $2.70\ \Omega$ resistor. Find the current and the terminal voltage.

21.19 ★★ *[Graph-reading.]* The graph shows the terminal voltage of a battery against the current it delivers.



•Figure description
A graph of terminal voltage in volts, from 0 to 7, against current in amperes, from 0 to 3. The data form a straight line starting at 6.0 volts at zero current and falling to 4.0 volts at 2.0 amperes, continuing to 3.0 volts at 3.0 amperes.

a. Find the emf and the internal resistance.
b. What current would flow if the terminals were shorted?
c. What load resistance would draw $1.5\ \text{A}$?

21.20 ★★ When a battery delivers $2.0\ \text{A}$, its terminal voltage is $11.4\ \text{V}$; when it delivers $5.0\ \text{A}$, its terminal voltage is $10.5\ \text{V}$. Find its emf and internal resistance.

21.21 ★★ Four $1.5\ \text{V}$ cells, each with an internal resistance of $0.20\ \Omega$, are connected in series to light a $4.0\ \Omega$ bulb. Find the current, the power delivered to the bulb, and the power wasted inside the cells.

21.22 ★★★ A battery of emf $\mathcal{E}$ and internal resistance r drives a load R. (a) Show that the efficiency, the fraction of the battery's output delivered to the load, is $R/(R+r)$. (b) Sketch the power to the load and the efficiency as functions of R/r on the same axes. (c) Explain why a power utility operates far from the condition of maximum power transfer.

Section 21.5 — RC Circuits

21.23 ★ *[Conceptual.]* Just after the switch is closed in a charging RC circuit, the capacitor acts like a plain wire; long afterward, it acts like a break in the circuit. Explain both statements using $v = q/C$.

Section 21.5.1 — Charging a Capacitor

21.24 ★ A $220\ \mu\text{F}$ capacitor is charged through a $4.7\ \text{k}\Omega$ resistor. Find the time constant and the time needed to reach 95% of the final charge.

21.25 ★★ An uncharged $10\ \mu\text{F}$ capacitor is charged through a resistor R by a $12\ \text{V}$ battery. Two seconds after the switch is closed, the voltage across the capacitor is $8.0\ \text{V}$. Find R.

21.26 ★★ *[Real-world: timing circuits.]* In an intermittent windshield wiper, a $100\ \mu\text{F}$ capacitor charges from a $9.0\ \text{V}$ supply through a resistor, and the wiper sweeps when the capacitor's voltage reaches $5.0\ \text{V}$. What resistance gives a $4.0\ \text{s}$ delay?

21.27 ★★ Show that a charging capacitor holds a quarter of its final energy at the moment $t = RC\ln 2$.

Section 21.5.2 — Discharging a Capacitor

21.28 ★ A $1000\ \mu\text{F}$ capacitor charged to $12\ \text{V}$ discharges through a $100\ \Omega$ resistor. Find the initial current, the time constant, and the charge remaining after $0.25\ \text{s}$.

21.29 ★★ *[Graph-reading.]* The graph shows the voltage across a capacitor as it discharges through a resistor.



•Figure description
A graph of capacitor voltage in volts, from 0 to 10, against time in seconds, from 0 to 8. The curve starts at 10 volts at time zero and decays exponentially, passing about 3.7 volts at 2 seconds, about 1.4 volts at 4 seconds, and approaching zero by 8 seconds.

a. Estimate the time constant from the graph, in two ways: from the time for the voltage to fall to 36.8% of its initial value, and from its half-life.
b. If the capacitance is $50\ \mu\text{F}$, what is the resistance?
c. What was the initial current?

21.30 ★★★ *[Real-world: electronics.]* No insulator is perfect, so a charged capacitor slowly discharges through its own dielectric. A parallel-plate capacitor is filled with a dielectric of dielectric constant κ and resistivity ρ. (a) Show that its self-discharge time constant is $\tau = \kappa\epsilon_0\rho$, independent of the plates' area and spacing. (b) Evaluate τ for glass, with $\kappa = 5.0$ and $\rho = 1.0 \times 10^{12}\ \Omega \cdot \text{m}$.

21.31 ★★★ A $12\ \text{V}$ battery is connected in series with a $4.0\ \Omega$ resistor R_1 and a parallel pair consisting of an $8.0\ \Omega$ resistor R_2 and an uncharged $5.0\ \mu\text{F}$ capacitor. (a) What is the current through the battery just after the circuit is connected? (b) What is it long afterward, and what is the charge on the capacitor then? (c) The battery is then disconnected, leaving the capacitor connected across R_2 alone. Find the time constant of the discharge.

21.32 ★★★ Kirchhoff's rules can be checked against the result for resistors in parallel. Two resistors R_1 and R_2 in parallel are connected to a battery of emf $\mathcal{E}$ and internal resistance r. Use the junction and loop rules, without using equation (21.6), to find the current through each resistor, and show that your result agrees with replacing the pair by $R_1R_2/(R_1 + R_2)$.

This page titled 21.E: DC Circuits (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22: THE NATURE OF MAGNETIC FIELDS

A compass needle swings to point north. A refrigerator magnet clings to the door. Deep inside a hospital's MRI scanner, a field tens of thousands of times stronger than Earth's lines up the nuclei in a patient's body, and in a ring 27 kilometers around beneath the border of France and Switzerland, more than a thousand magnets steer protons moving at 99.999999% of the speed of light. All of these are magnetic effects, and all of them come down to one idea: a moving charge feels a force in a magnetic field, a force that pushes it *sideways*, at right angles to its motion.

That sideways force makes magnetism very different from electricity. An electric field pushes a charge along the field, whether it moves or not. A magnetic field ignores a charge at rest, and pushes a moving one in a direction perpendicular both to its velocity and to the field. This chapter describes that force and follows its consequences, from charged particles spiraling in circles to the discovery of the electron, the particle accelerators of modern physics, the forces and torques that turn every electric motor, and the Hall effect that lets a phone sense which way it is pointing. Where magnetic fields come from, and in particular how currents produce them, is the subject of Chapter 23.

Learning Objectives

- Describe magnetic fields and field lines, and find the magnetic force on a moving charge from $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$.
- Analyze the circular and helical motion of a charged particle in a uniform magnetic field, and apply the Lorentz force in combined electric and magnetic fields.
- Explain how magnetic fields were used to discover the electron and to accelerate particles in cyclotrons and synchrotrons.
- Calculate the magnetic force on a current-carrying wire and the torque on a current loop, and use the magnetic dipole moment.
- Explain the Hall effect and use it to find the sign and density of charge carriers.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

22.1: Magnetic Fields
22.2: Motion of a Charged Particle in Magnetic Fields
22.3: Applications of Magnetic Fields
22.S: The Nature of Magnetic Fields (Summary)
22.E: The Nature of Magnetic Fields (Problems)

This page titled 22: The Nature of Magnetic Fields is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22.1: Magnetic Fields

Magnetism was known long before electricity was understood. The ancient Greeks, and the Chinese even earlier, knew that certain stones, lodestones of the mineral magnetite, attract iron, and by the eleventh century Chinese navigators were steering by compasses made of magnetized needles. Every magnet has two ends, or **poles**, where its effects are strongest. A freely pivoted magnet turns so that one pole points roughly north; that end is called its north pole (N), and the other its south pole (S). Like poles repel each other and unlike poles attract, much like electric charges.

The resemblance ends there. Electric charges can be separated: a body can carry a net positive or negative charge. Magnetic poles cannot. Cut a bar magnet in half, and each half is a complete magnet with its own north and south poles; cut it again, and the same thing happens, all the way down to single atoms. No isolated north or south pole, no *magnetic monopole*, has ever been found, despite careful searches (Section 23.6).

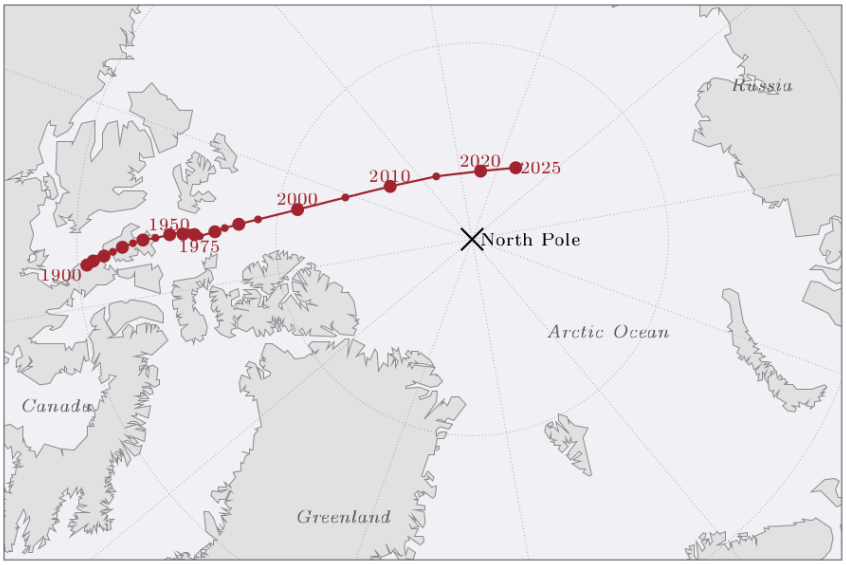


Figure 22.1. Drift of Earth's north magnetic pole (the spot where a compass needle would point straight down) from 1900 to 2025, computed from the International Geomagnetic Reference Field (IGRF-14). Small dots mark every five years and large dots every ten. For most of the twentieth century the pole crept through the Canadian Arctic at roughly 5 to 15 kilometers a year; in the 1990s it sped up to 55 to 60 kilometers a year, and it is now slowing down (about 35 kilometers a year) as it heads toward Siberia.

•Figure description

A map of the Arctic seen from above the geographic North Pole, which is marked with an X. Canada and its Arctic islands are at the left, Greenland at the bottom, and Russia at the upper right, with the Arctic Ocean between them. A red line of dots shows the position of the north magnetic pole from 1900 to 2025. From 1900 to about 1975 the dots are bunched together among the islands of northern Canada, showing slow motion. After 1990 they spread far apart as the pole races across the Arctic Ocean, passing near the geographic North Pole around 2015 to 2020 and heading toward Russia by 2025.

Earth itself is a giant magnet. A compass needle's north pole points toward the geographic north, so, since unlike poles attract, the magnetic pole near Earth's geographic North Pole is actually a magnetic *south* pole. It is not fixed: it drifts around 5–60 kilometers each year (Figure 22.1), and over geological time the whole field has reversed direction hundreds of times, with reversals recorded in the magnetized rocks of the ocean floor.

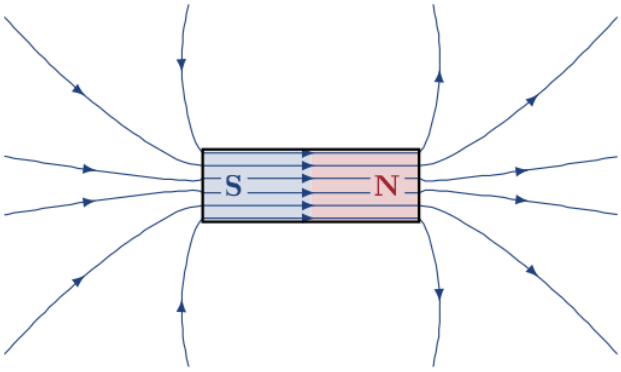


Figure 22.2. Magnetic field lines of a bar magnet, computed. Outside the magnet they run from the north pole around to the south pole; inside it they return from south to north. Every line closes on itself. The same pattern can be seen by sprinkling iron filings on a sheet of paper over a magnet.

•Figure description

The magnetic field lines of a bar magnet, computed. A horizontal rectangular magnet is shaded red on its right half, labeled N, and blue on its left half, labeled S. Field lines leave the N end, fan out, and curve around through the space above and below the magnet; the lines from the edges of the face turn back sharply, while those from near the center of the face run far out of the frame. Lines come back in to the S end from all directions. Inside the magnet, the lines run straight from the S end to the N end, so that every line is a closed loop. Arrowheads show the direction of the field: away from N outside the magnet, and from S to N inside it.

The magnetic field. As with electric forces, it is most useful to describe magnetic forces through a field. A magnet, or a current, sets up a **magnetic field** $\mathbf{B}$ in the space around it, and that field exerts forces on other magnets and moving charges. The direction of $\mathbf{B}$ at any point is the direction in which the north pole of a small compass needle points at that location. Like the electric field, the magnetic field can be drawn with **magnetic field lines**, tangent to $\mathbf{B}$ at every point and crowded together where the field is strong (Figure 22.2). Outside a magnet, they run from its north pole to its south pole. But they do not end there. They continue through the magnet, from the south pole back to the north, and close on themselves. Magnetic field lines never begin or end, because there are no magnetic charges for them to begin or end on. This is the content of Gauss's law for magnetism, $\oint \mathbf{B} \cdot d\mathbf{A} = 0$, which Section 23.6 makes precise.

The magnetic force on a moving charge. In 1820 the Danish physicist Hans Christian Oersted noticed that a compass needle near a wire would turn whenever a current flowed in the wire. Magnetism, it turned out, is produced by moving charges, and in turn it acts on moving charges. The experimental facts about the force on a charge q moving with velocity $\mathbf{v}$ through a magnetic field are these:

- A charge at rest feels no magnetic force. The force is proportional to the speed v, and to the charge q.
- A charge moving parallel (or antiparallel) to $\mathbf{B}$ feels no force. For other directions, the force is proportional to $\sin\theta$, where θ is the angle between $\mathbf{v}$ and $\mathbf{B}$.
- The force is perpendicular both to $\mathbf{v}$ and to $\mathbf{B}$, and it reverses if the sign of q is reversed.

All of this is summarized by the cross product of Chapter 3, and it defines the magnetic field:

Key Equation: Magnetic force on a moving charge

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}, \qquad F_B = |q|vB\sin\theta. \tag{22.1}$$

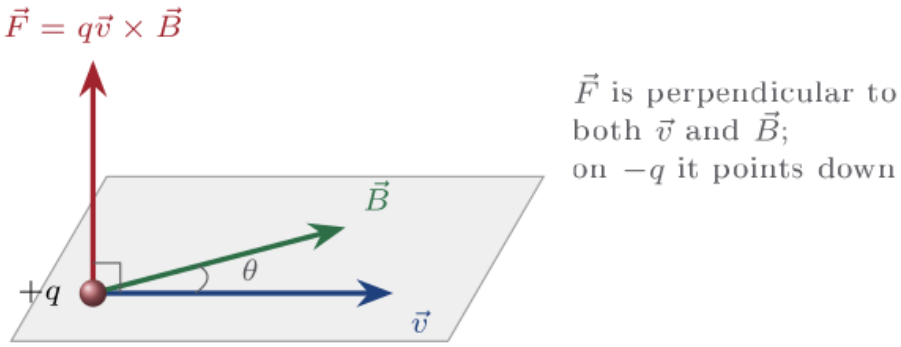


Figure 22.3. The magnetic force on a positive charge is perpendicular to the plane containing $\mathbf{v}$ and $\mathbf{B}$, in the direction given by the right-hand rule for $\mathbf{v} \times \mathbf{B}$. Its magnitude is $|q|vB\sin\theta$.

•Figure description

A three-dimensional sketch of the magnetic force on a positive charge q at the origin. The velocity v points to the right, and the magnetic field B points up and to the right, at an angle theta from v, both lying in a lightly shaded horizontal plane. The force F points straight up, perpendicular to the plane containing v and B, and is labeled F equals q v cross B. A note says that on a negative charge the force points down.

The direction follows from the right-hand rule for cross products (Figure 22.3): point the fingers of your right hand along $\mathbf{v}$, curl them toward $\mathbf{B}$, and your thumb points along $\mathbf{v} \times \mathbf{B}$. That is the direction of the force on a positive charge; on a negative charge the force points the opposite way.

Units. From equation (22.1), the SI unit of magnetic field is a newton per coulomb per meter per second. It is called the **tesla** (T), after the inventor Nikola Tesla:

$$1\,\mathrm{T} = 1\,\frac{\mathrm{N}}{\mathrm{C}\cdot\mathrm{m/s}} = 1\,\frac{\mathrm{N}}{\mathrm{A}\cdot\mathrm{m}}. \tag{22.1.1}$$

A single tesla is a large field. Earth's field at the surface is about $50\ \mu\mathrm{T}$, and an older unit, the **gauss** (G), with $1\ \mathrm{G} = 10^{-4}\ \mathrm{T}$, is still widely used for small fields; Earth's field is about half a gauss. Table 22.1 lists some typical values.

Table 22.1. Some typical magnetic fields.

Source	Magnetic field (T)
Human brain (measured outside the skull)	10^{-13} to 10^{-12}
Interstellar space	10^{-10}
Earth's surface	5×10^{-5}
Refrigerator magnet	5×10^{-3}
Neodymium magnet, at its surface	0.5 to 1.4
MRI scanner	1.5 to 7
Large Hadron Collider bending magnets	8.3
Strongest steady laboratory field	45
Surface of a neutron star	10^{8}
Surface of a magnetar	10^{10} to 10^{11}

Drawing fields in three dimensions. Magnetic forces are perpendicular to the field, so problems are unavoidably three-dimensional. A field pointing *into* the page is drawn as a grid of crosses, ⊗, like the tail feathers of an arrow flying away from you; a field pointing *out of* the page is drawn as dots, ⊙, like the point of an arrow coming toward you. Both appear in the figures that follow.

Example 22.1: The magnetic force on a proton

A proton moves at $4.0 \times 10^6\ \text{m/s}$ at right angles to Earth's magnetic field, $5.0 \times 10^{-5}\ \text{T}$. Find the magnetic force on it, and compare it with the proton's weight.

Solution

With $\theta = 90^\circ$,

$$F_B = evB = (1.602 \times 10^{-19})(4.0 \times 10^6)(5.0 \times 10^{-5}) = \boxed{3.2 \times 10^{-17}\,\text{N}}. \tag{22.1.2}$$

The proton's weight is $mg = (1.67 \times 10^{-27})(9.8) = 1.6 \times 10^{-26}\ \text{N}$, so the magnetic force is about two billion times larger. A field that barely turns a compass needle dominates the motion of a fast charged particle.

Example 22.2: Finding the force vector

An electron moves with velocity $\mathbf{v} = (3.0\ \hat{\imath} + 4.0\ \hat{\jmath}) \times 10^6\ \text{m/s}$ through a uniform field $\mathbf{B} = (0.020\ \text{T})\ \hat{\imath}$. Find the magnetic force on it.

Solution

First the cross product, using $\hat{\imath} \times \hat{\imath} = 0$ and $\hat{\jmath} \times \hat{\imath} = -\hat{k}$:

$$\mathbf{v} \times \mathbf{B} = (10^6)(0.020)\left[(3.0)\,\hat{\imath} \times \hat{\imath} + (4.0)\,\hat{\jmath} \times \hat{\imath}\right] = -(8.0 \times 10^4\,\text{T} \cdot \text{m/s})\,\hat{k}. \tag{22.1.3}$$

Then, with $q = -e$,

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B} = (-1.602 \times 10^{-19})(-8.0 \times 10^4)\,\hat{k} = \boxed{(1.3 \times 10^{-14}\,\text{N})\,\hat{k}}. \tag{22.1.4}$$

Only the component of $\mathbf{v}$ perpendicular to $\mathbf{B}$, the $\hat{\jmath}$ part, contributes. As a check on the magnitude, $v = 5.0 \times 10^6\ \text{m/s}$ and $\sin\theta = 4.0/5.0$, so $evB\sin\theta = (1.602 \times 10^{-19})(5.0 \times 10^6)(0.020)(0.80) = 1.3 \times 10^{-14}\ \text{N}$.

Magnetic forces do no work. Because $\mathbf{F}_B$ is always perpendicular to $\mathbf{v}$, it is always perpendicular to the displacement of the charge, and it does no work on it (Chapter 7). A magnetic field can change the *direction* of a charged particle's velocity, but never its speed or its kinetic energy. To speed a particle up, you need an electric field.

Checkpoint 22.1

A positive charge moves to the right, in the plane of the page, through a magnetic field that points into the page. (a) Which way is the force on it? (b) Which way would the force be on an electron moving in the same direction? (c) What if the positive charge moved into the page, along the field?

Answer

Answers to Checkpoints

Checkpoint 22.1. Use $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ with $\mathbf{v}$ to the right and $\mathbf{B}$ into the page. (a) Point your fingers right, curl them into the page, and your thumb points *up the page*: that is the force on the positive charge. (b) An electron's charge is negative, so its force is reversed: down the page. (c) If $\mathbf{v}$ is parallel to $\mathbf{B}$, then $\mathbf{v} \times \mathbf{B} = 0$ and there is no magnetic force at all.

This page titled 22.1: Magnetic Fields is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22.2: Motion of a Charged Particle in Magnetic Fields

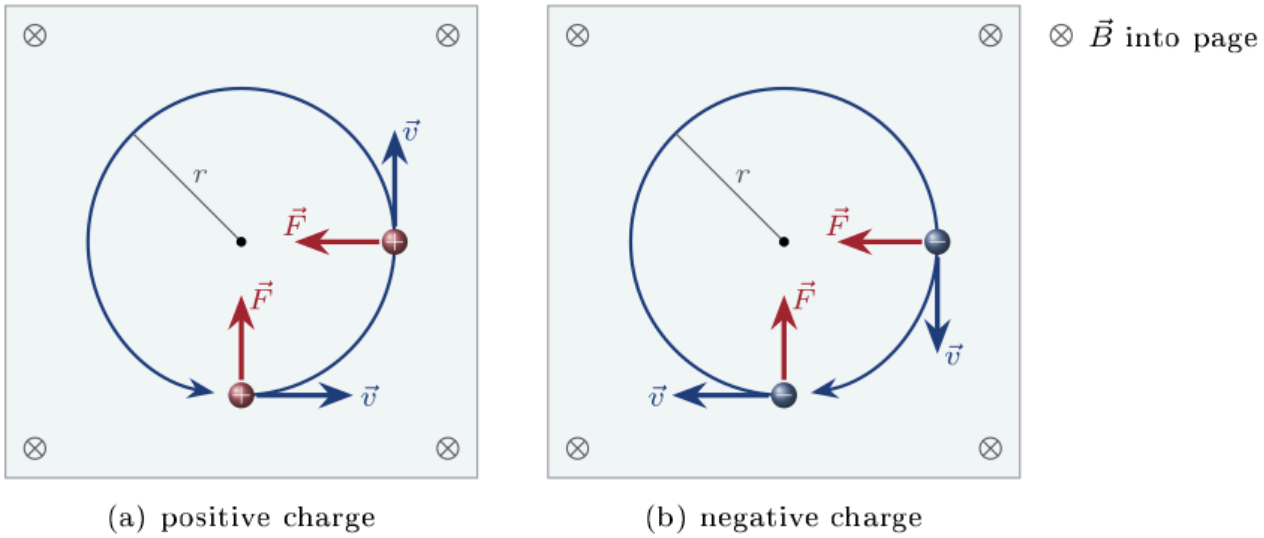


Figure 22.4. A charge moving perpendicular to a uniform magnetic field (into the page) travels in a circle, and the magnetic force always points toward the center. (a) A positive charge circulates counterclockwise. (b) A negative charge circulates the other way, clockwise.

•Figure description

Two panels, each a lightly shaded square region of uniform magnetic field pointing into the page, marked by a cross in a circle at each corner. Panel a: a positive charge moves in a circle of radius r. At the bottom of the circle it moves to the right and the magnetic force on it points up, toward the center; at the right side it moves up and the force points left, toward the center. It circulates counterclockwise. Panel b: a negative charge in the same field. At the bottom of the circle it moves to the left and the force points up, toward the center; at the right side it moves down and the force points left, toward the center. It circulates clockwise.

Consider a particle of charge q and mass m moving with velocity v perpendicular to a uniform magnetic field (Figure 22.4). The force has constant magnitude $|q|vB$, since the speed never changes, and it is always perpendicular to the velocity. A force of constant size that always acts at right angles to the motion is exactly what produces uniform circular motion (Section 6.3). The magnetic force supplies the centripetal force:

$$|q|vB = \frac{mv^2}{r}. \tag{22.2.1}$$

The particle moves in a circle of radius

Key Equation: Radius of the circular orbit

$$r = \frac{mv}{|q|B} = \frac{p}{|q|B}. \tag{22.2}$$

Faster or heavier particles move in larger circles; stronger fields bend them more tightly. The time for one orbit is $T = 2\pi r/v = 2\pi m/|q|B$, and the angular frequency is

Key Equation: Cyclotron frequency

$$\omega = \frac{|q|B}{m}, \qquad f = \frac{|q|B}{2\pi m}. \tag{22.3}$$

This is the **cyclotron frequency**. Remarkably, it does not depend on the particle's speed or on the size of its orbit. A fast particle travels a larger circle, but it goes around it in exactly the same time as a slow one. That fact is what makes the cyclotron of Section

22.3 possible.

Example 22.3: An electron in a magnetic field

An electron moving at $2.0 \times 10^6 \text{ m/s}$ enters a uniform field of 1.0 mT, moving perpendicular to it. Find the radius of its orbit and the time for one revolution.

Solution

$$r = \frac{m_e v}{eB} = \frac{(9.11 \times 10^{-31})(2.0 \times 10^6)}{(1.602 \times 10^{-19})(1.0 \times 10^{-3})} = \boxed{1.1 \text{ cm}}, \tag{22.2.2}$$

and

$$T = \frac{2\pi m_e}{eB} = \frac{2\pi(9.11 \times 10^{-31})}{(1.602 \times 10^{-19})(1.0 \times 10^{-3})} = \boxed{36 \text{ ns}}. \tag{22.2.3}$$

The electron circles about 28 million times a second, whatever its speed, as long as it is nonrelativistic.

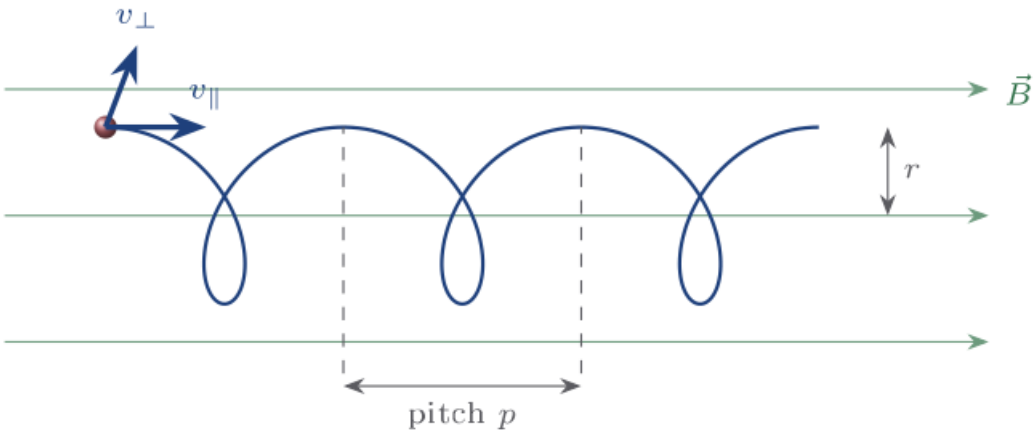


Figure 22.5. A particle whose velocity has a component along $\mathbf{B}$ moves in a helix. The circular part of the motion has radius $r = mv_\perp/|q|B$, and the pitch is $p = v_\parallel T = 2\pi m v_\parallel/|q|B$.

•Figure description

A charged particle spiraling in a helix around uniform magnetic field lines that point to the right. The particle's velocity has a component v parallel along the field and a component v perpendicular across it. The helix has radius r, and the distance along the field between successive turns, the pitch, is marked p.

Helical motion. If the velocity has a component along the field, that component feels no force and stays constant. The perpendicular component still produces circular motion. The combination is a **helix**, a spiral wound around the field lines (Figure 22.5). The radius comes from the perpendicular component, $r = mv_\perp/|q|B$; the particle advances along the field by a distance $p = v_\parallel T$, the pitch of the helix, in each revolution.

Example 22.4: A proton in a helix

A proton moves at $1.0 \times 10^6 \text{ m/s}$ at 30° to a uniform field of 0.10 T. Find the radius and the pitch of its helical path.

Solution

The velocity components are $v_\perp = v \sin 30^\circ = 5.0 \times 10^5 \text{ m/s}$ and $v_\parallel = v \cos 30^\circ = 8.7 \times 10^5 \text{ m/s}$. The radius is

$$r = \frac{mv_\perp}{eB} = \frac{(1.67 \times 10^{-27})(5.0 \times 10^5)}{(1.602 \times 10^{-19})(0.10)} = \boxed{5.2 \text{ cm}}. \tag{22.2.4}$$

The period is $T = 2\pi m/eB = 6.6 \times 10^{-7}$ s, so the pitch is

$$p = v_{\parallel} T = (8.7 \times 10^{5})(6.6 \times 10^{-7}) = \boxed{57\text{ cm}}. \tag{22.2.5}$$

The proton makes a loose spiral, advancing more than half a meter along the field for each turn of radius five centimeters.

Note: The aurora

Charged particles from the Sun, streaming out in the solar wind, are captured by Earth's magnetic field and spiral along its field lines. Where the field lines converge near the magnetic poles, the growing field reflects many of the particles back, so they bounce between the hemispheres, trapped in the Van Allen radiation belts. Others follow the field lines down into the upper atmosphere near the poles, where they collide with oxygen and nitrogen atoms and make them glow: the green and red curtains of the aurora borealis and aurora australis. The same principle, a *magnetic bottle*, is used to confine hot plasmas in fusion research (Chapter 43).

The Lorentz Force

When both an electric field and a magnetic field are present, a charged particle feels both forces. The total electromagnetic force on a charge q moving with velocity $\mathbf{v}$ is the **Lorentz force**, named for the Dutch physicist Hendrik Lorentz:

Key Equation: The Lorentz force

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right). \tag{22.4}$$

Together with Maxwell's equations, which say how the fields are produced (Chapter 27), this one equation contains everything that electric and magnetic fields do to charged matter.

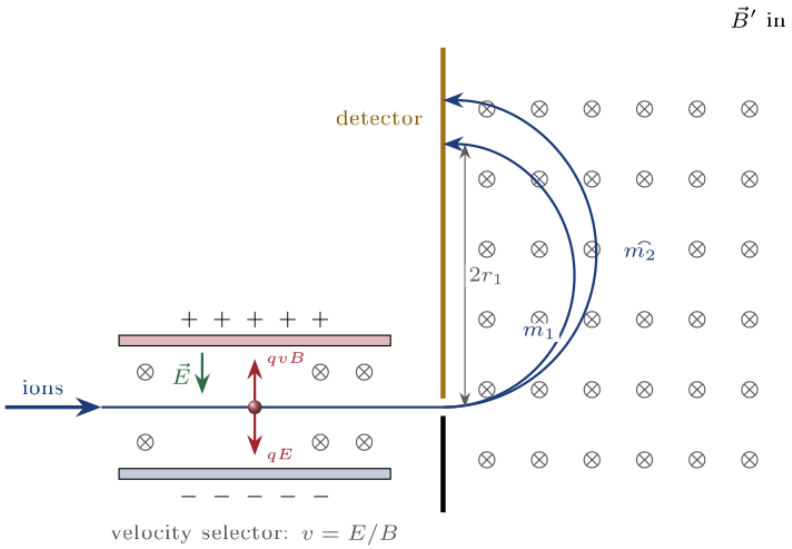


Figure 22.6. A velocity selector passes only ions with speed $v = E/B$, for which the electric and magnetic forces cancel. In the mass spectrometer beyond it, the ions move in semicircles of radius $r = mv/qB'$, so ions of different mass land at different places on the detector.

•Figure description

A velocity selector followed by a mass spectrometer. On the left, a horizontal beam of positive ions enters between two horizontal parallel plates, the top one positive and the bottom one negative, so the electric field E points down. A magnetic field B points into the page in the same region. On an ion moving right, the electric force q E points down and the magnetic force q v B points up; they balance for ions with speed E over B, which pass straight through a slit. Beyond the slit is a region with only a magnetic field B prime, into the page. There the ions travel in semicircles, curving upward and back to the left, and strike a vertical detector plate that rises above the slit. Two semicircles are shown, a smaller one for the lighter ions of mass m sub 1 and a larger one for heavier ions of mass m sub 2, landing at different heights on the detector.

The velocity selector. Arrange a uniform electric field and a uniform magnetic field perpendicular to each other and to a beam of charged particles (Figure 22.6). For a positive particle moving to the right, the electric force $q\mathbf{E}$ points down and the magnetic force $q\mathbf{v} \times \mathbf{B}$ points up. The two cancel if $qE = qvB$, that is, for particles with exactly the speed

$$v = \frac{E}{B}. \tag{22.5}$$

Those particles pass straight through, undeflected, and emerge through a slit. Faster particles are pushed up by the larger magnetic force, and slower ones down, and they miss the slit. This **velocity selector** picks out particles of a single speed, whatever their charge or mass.

The mass spectrometer. Send the selected particles into a second region containing only a uniform magnetic field B'. There they travel in semicircles of radius $r = mv/qB'$, and land on a detector at a distance $2r$ from the entrance slit. Since v, q, and B' are the same for all of them, the radius is proportional to the mass. This is a **mass spectrometer**, which sorts atoms and molecules by mass with great precision. Mass spectrometers were used to discover isotopes, to separate the uranium-235 for the first atomic bomb, and today to identify molecules in everything from drug tests to the atmospheres of other planets.

Example 22.5: Separating isotopes of carbon

Singly charged ions of carbon-12 ($m = 12.0$ u) and carbon-14 ($m = 14.0$ u) pass through a velocity selector with $E = 1.0 \times 10^5$ V/m and $B = 0.10$ T, and then enter a region where $B' = 0.50$ T. How far apart do they land on the detector? (1 u $= 1.661 \times 10^{-27}$ kg.)

Solution

The selected speed is $v = E/B = 1.0 \times 10^6$ m/s for both isotopes. The radius for carbon-12 is

$$r_{12} = \frac{mv}{eB'} = \frac{(12.0)(1.661 \times 10^{-27})(1.0 \times 10^6)}{(1.602 \times 10^{-19})(0.50)} = 0.249\,\text{m}, \tag{22.2.6}$$

and for carbon-14, heavier by a factor $14/12$, $r_{14} = 0.290$ m. Each lands at a distance $2r$ from the slit, so they are separated by

$$2(r_{14} - r_{12}) = 2(0.290 - 0.249) = \boxed{8.3\,\text{cm}}, \tag{22.2.7}$$

easily resolved. Accelerator mass spectrometers built on this principle can count individual carbon-14 atoms, making radiocarbon dating possible with samples of less than a milligram (Chapter 43).

Checkpoint 22.2

In a velocity selector, (a) what happens to a particle moving faster than E/B? (b) Does the selected speed depend on the sign of the charge? Check by reversing the charge in Figure 22.6.

Answer

Answers to Checkpoints

Checkpoint 22.2. (a) For a faster particle the magnetic force, qvB, exceeds the electric force, qE, which does not depend on speed. The two no longer cancel, so the particle is deflected in the direction of the magnetic force and misses the exit slit. (b) No. Reversing the charge reverses *both* forces, so at $v = E/B$ they still cancel. The selector passes particles of that speed whatever their charge or mass.

This page titled 22.2: Motion of a Charged Particle in Magnetic Fields is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22.3: Applications of Magnetic Fields

The magnetic force on moving charges is at work in instruments that changed the history of physics and in devices found in every home. This section describes five of them.

Discovery of the Electron

By the late nineteenth century, physicists knew that a glass tube pumped nearly empty of air, with a high voltage between two metal electrodes sealed inside it, glowed where something emitted by the negative electrode, the cathode, struck the far end of the tube. These **cathode rays** traveled in straight lines and cast shadows, and they could be bent by a magnet. But what were they? Some physicists thought they were a kind of wave in the ether; others, that they were particles.

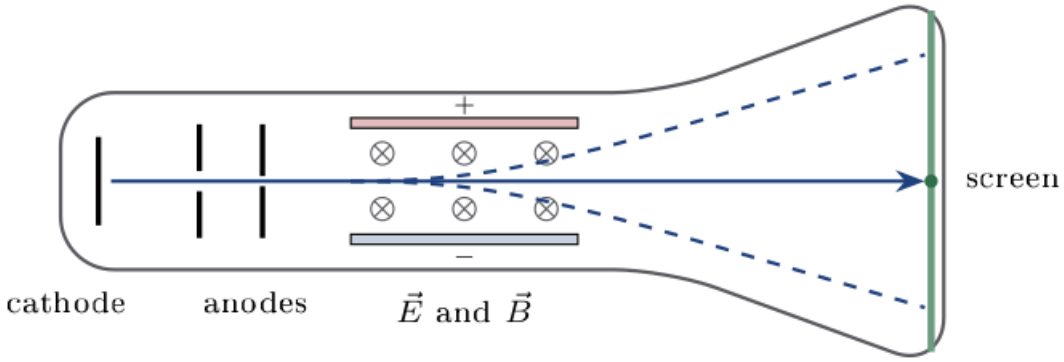


Figure 22.7. Thomson's apparatus (schematic). Electrons from the cathode are accelerated and collimated by the anodes, then pass through crossed electric and magnetic fields. When the fields are balanced the beam goes straight; with one field alone it is deflected (dashed).

•Figure description

A schematic of Thomson's cathode-ray tube, drawn as a long evacuated glass tube. At the left end is the cathode; to its right, the anode, a plate with a hole, accelerates the rays, and a second plate with a slit narrows them into a beam. In the middle of the tube are two horizontal deflecting plates, the upper one positive, producing a downward electric field; the same region is filled with a magnetic field into the page, shown by crosses in circles. At the right end is a fluorescent screen. The undeflected beam travels straight through to the center of the screen; dashed paths show the beam bent up or down when only one of the fields is on.

In 1897 J. J. Thomson settled the question at the Cavendish Laboratory in Cambridge with the apparatus sketched in Figure 22.7. The rays from the cathode were accelerated by an anode and passed through slits to form a narrow beam. The beam then passed between two deflecting plates, which could apply an electric field, and through a magnetic field at right angles to it, before striking a fluorescent screen. Thomson measured the beam's deflection by the electric field alone (as in Figure 16.13), and then adjusted the magnetic field until the beam went straight through: the arrangement is a velocity selector, and it gave the speed of the rays, $v = E/B$. With the magnetic field alone, the beam curves in a circle of radius $r = mv/qB$. Substituting the speed,

$$\frac{q}{m} = \frac{v}{rB} = \frac{E}{rB^2}. \tag{22.6}$$

Every quantity on the right can be measured. Thomson found that cathode rays have a definite charge-to-mass ratio (Chapter 16) that is the same no matter the metal the cathode was made of or the gas that remained in the tube. The rays were particles, negatively charged, present in all matter. And their q/m was more than a thousand times larger than that of hydrogen ions, the lightest particles then known: either their charge was very large or their mass was very small. When Millikan measured the elementary charge a decade later (Example 16.12), the answer became clear. The particles, soon called **electrons**, have the same size of charge as a hydrogen ion and about $1/1836$ of its mass. The atom, long assumed to be indivisible, had parts (Chapter 41).

Example 22.6: Measuring e/m

In a Thomson-type experiment, an electric field of $1.5 \times 10^4\ \mathrm{V/m}$ is balanced by a magnetic field of $0.50\ \mathrm{mT}$. With the electric field switched off, the beam bends in a circle of radius $34\ \mathrm{cm}$. Find the speed of the electrons and their charge-to-mass ratio.

Solution

The speed is $v = E/B = (1.5 \times 10^4)/(0.50 \times 10^{-3}) = \boxed{3.0 \times 10^7\ \mathrm{m/s}}$, a tenth of the speed of light. Then

$$\frac{e}{m} = \frac{E}{rB^2} = \frac{1.5 \times 10^4}{(0.34)(0.50 \times 10^{-3})^2} = \boxed{1.8 \times 10^{11}\ \mathrm{C/kg}}, \tag{22.3.1}$$

in agreement with the modern value, $1.76 \times 10^{11}\ \mathrm{C/kg}$. The accelerating voltage that gave the electrons this speed follows from $eV = \frac{1}{2}mv^2$: about $2.6\ \mathrm{kV}$.

Cyclotrons and Synchrotrons

To probe the atomic nucleus, physicists needed beams of particles with energies of millions of electron volts. Producing such energies with a single voltage is very difficult. In 1931 Ernest Lawrence found a way around the problem: give the particles many small kicks instead of one big one, and use a magnetic field to bring them back for each kick.

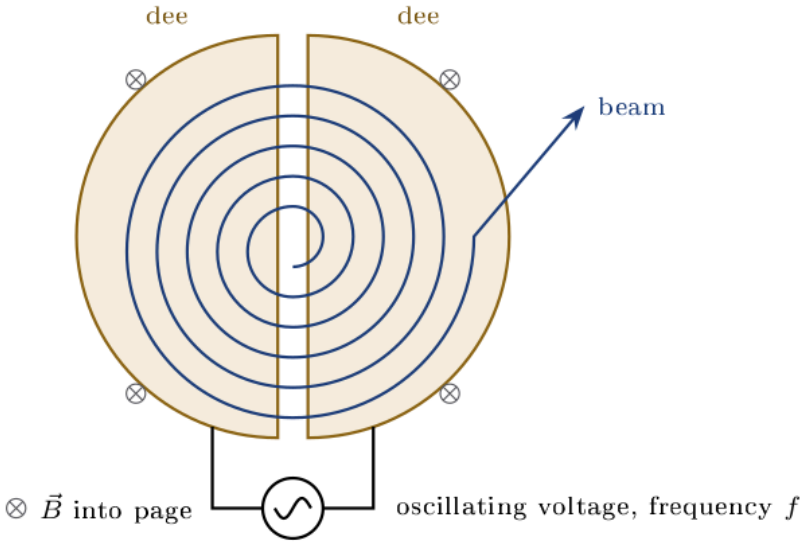


Figure 22.8. A cyclotron, seen from above. The magnetic field bends the particles into semicircles inside the dees, and the oscillating voltage accelerates them each time they cross the gap. Because the orbital period does not depend on speed, one fixed frequency keeps them in step.

•Figure description

A top view of a cyclotron. Two D-shaped hollow electrodes, the dees, face each other across a narrow gap, forming a circle. A uniform magnetic field points into the page. An oscillating voltage source, drawn below as a circle with a wavy line, is connected to both dees. A particle released near the center follows a spiral path made of semicircles of steadily increasing radius, crossing the gap twice per revolution and gaining energy at each crossing, until it leaves the edge of the dees as the output beam.

A **cyclotron** consists of two hollow, D-shaped metal electrodes (called “dees”), in a vacuum chamber between the poles of a large electromagnet (Figure 22.8). An oscillating voltage is applied between the dees. Inside each dee there is no electric field, and a charged particle simply travels in a semicircle under the magnetic force. Each time it crosses the gap between the dees, the voltage has reversed, and the particle is accelerated again. With each crossing its speed increases and it moves out to a semicircle of larger radius. But by equation (22.3), the time for each semicircle is the same, so if the voltage oscillates at the cyclotron frequency $f = qB/2\pi m$, the particle arrives at the gap in step with the voltage every time. The particle spirals outward until it reaches the edge of the dees at radius R, with speed $v = qBR/m$ and kinetic energy

$$K = \frac{1}{2}mv^2 = \frac{q^2B^2R^2}{2m}. \tag{22.7}$$

Example 22.7: A proton cyclotron

A cyclotron with dees of radius $0.50\ \text{m}$ in a field of $1.5\ \text{T}$ accelerates protons. Find (a) the frequency of the oscillating voltage, (b) the protons' final kinetic energy, and (c) the number of revolutions they make if they gain $50\ \text{keV}$ at each gap crossing.

Solution

(a) The cyclotron frequency is

$$f = \frac{eB}{2\pi m_p} = \frac{(1.602\times10^{-19})(1.5)}{2\pi(1.673\times10^{-27})} = \boxed{23\,\text{MHz}}, \tag{22.3.2}$$

a radio frequency.

(b) From equation (22.7),

$$K = \frac{e^2B^2R^2}{2m_p} = \frac{(1.602\times10^{-19})^2(1.5)^2(0.50)^2}{2(1.673\times10^{-27})} = 4.3\times10^{-12}\,\text{J} = \boxed{27\,\text{MeV}}. \tag{22.3.3}$$

(c) Each revolution includes two gap crossings, worth $100\ \text{keV}$, so reaching $27\ \text{MeV}$ takes about $\boxed{270}$ revolutions. A modest voltage, used hundreds of times, produces an energy that would need tens of millions of volts to reach in one step.

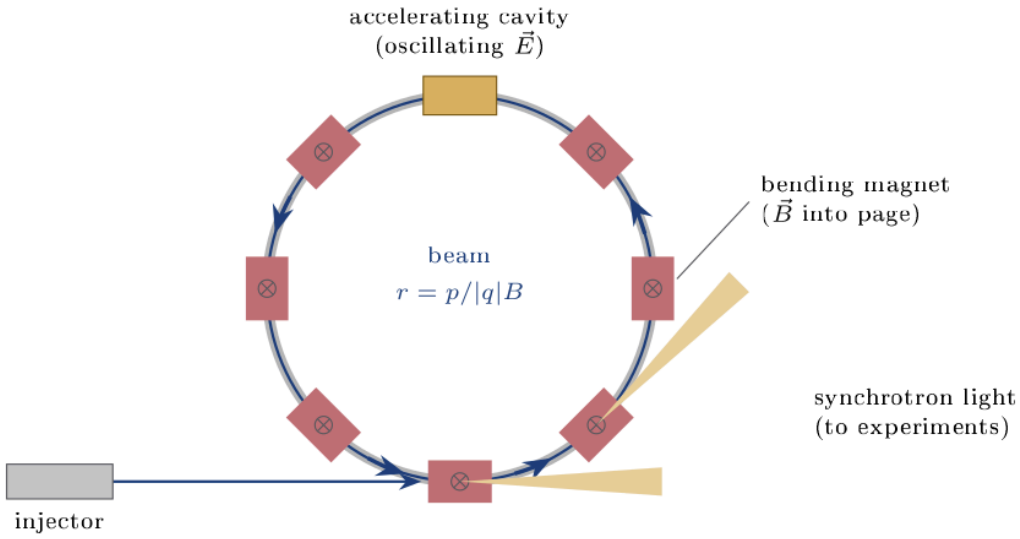


Figure 22.9. A synchrotron, seen from above (not to scale). Bending magnets hold the particles on a ring of fixed radius, and an accelerating cavity gives them a kick on every lap. As their momentum grows, the magnets' field is raised in step so that $r = p/|q|B$ stays the same. Electrons bent by the magnets give off synchrotron light along the direction they are moving, tangent to the ring.

•Figure description

A synchrotron seen from above, not to scale. A circular beam pipe carries the particles counterclockwise, shown by arrows along the ring. Seven bending magnets sit around the ring, each marked with a cross in a circle because its field points into the page. At the top of the ring is an accelerating cavity with an oscillating electric field. Particles enter from an injector at the lower left, joining the ring tangentially at the bottom. From two of the bending magnets at the lower right, pale wedges of synchrotron light leave tangent to the ring, toward experiments. The center of the ring is labeled beam, r equals p over the magnitude of q times B.

Cyclotrons are still used to make the short-lived radioactive isotopes for medical imaging and to deliver proton beams for cancer therapy. But they have a limit. The protons of Example 22.7 end with a speed of $7.2\times10^7\ \text{m/s}$, about a quarter of the speed of light, and at such speeds the relativistic mass effects of Chapter 35 lengthen the orbital period, so the particles fall out of step with the voltage. The cure is to change the field or the frequency as the particles speed up.

Synchrotrons. In a **synchrotron** the particles travel around a ring of fixed radius, a narrow evacuated pipe threading a sequence of bending magnets (Figure 22.9), and are accelerated by electric fields at a few points around the ring. As the particles gain energy, the magnetic field is increased in step, or in synchrony, to keep them on the same circle. Equation (22.2), written as $r = p/|q|B$, remains exactly correct at relativistic speeds provided p is the relativistic momentum γmv (equation (35.9)). The largest synchrotron is the Large Hadron Collider near Geneva.

Example 22.8: The Large Hadron Collider

The LHC is designed to bring protons to an energy of $7.0\ \mathrm{TeV}$ in bending magnets with a field of $8.3\ \mathrm{T}$. At this energy the proton's momentum is essentially $p = E/c$. What is the radius of curvature of the protons' path?

Solution

The momentum is

$$p = \frac{E}{c} = \frac{(7.0 \times 10^{12})(1.602 \times 10^{-19}\,\mathrm{J})}{3.00 \times 10^{8}\,\mathrm{m/s}} = 3.7 \times 10^{-15}\,\mathrm{kg \cdot m/s}, \tag{22.3.4}$$

so

$$r = \frac{p}{eB} = \frac{3.7 \times 10^{-15}}{(1.602 \times 10^{-19})(8.3)} = \boxed{2.8\,\mathrm{km}}. \tag{22.3.5}$$

The bending magnets, more than a thousand of them each $15\ \mathrm{m}$ long, fill about two-thirds of the 27-kilometer ring; the rest holds the accelerating cavities, focusing magnets, and detectors. The ring is so large because even the strongest practical magnets can bend such energetic protons only gently.

Note: Synchrotron radiation

A charged particle moving in a circle is accelerating, and accelerating charges radiate electromagnetic waves (Chapter 27). For protons this loss is small, but electrons, far lighter, radiate intensely when they are bent at high energy. This **synchrotron radiation** limits how much energy electron synchrotrons can reach, but it is also enormously useful: dedicated electron storage rings around the world, called light sources, produce intense beams of x-rays for studying the structure of proteins, materials, and ancient artifacts.

Magnetic Force on a Current-Carrying Wire

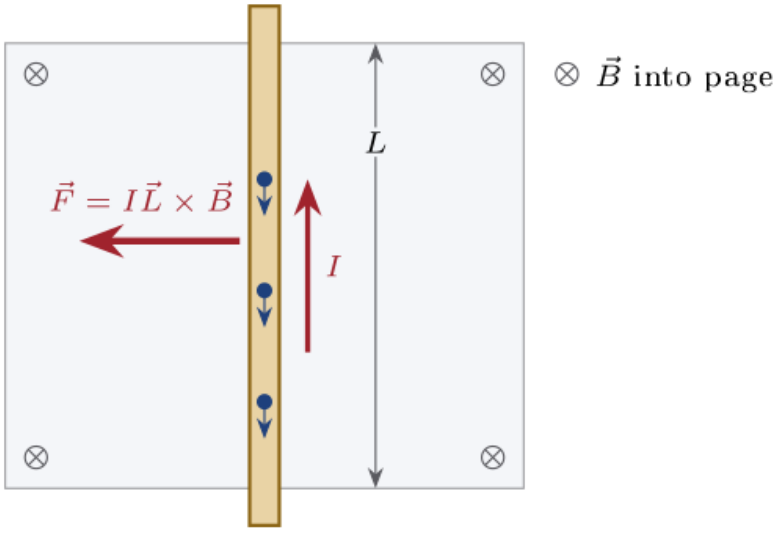


Figure 22.10. A wire carrying current I upward through a magnetic field directed into the page is pushed to the left. The electrons (blue) drift downward, and the magnetic force on each of them points left; they carry the wire with them.

•Figure description

A vertical straight wire of length L in a uniform magnetic field pointing into the page, shown as a lightly shaded region with a cross in a circle at each corner. The wire extends above and below the field region, and L is the length inside it. The current I flows upward. The magnetic force on the wire, F equals I L cross B, points to the left, perpendicular to both the wire and the field. Small electrons inside the wire drift downward, opposite to the current, and each feels a force to the left as well.

A current is a stream of moving charges, so a magnetic field exerts a force on a wire that carries a current. Consider a straight segment of wire of length L and cross-sectional area A, carrying current I in a uniform field $\mathbf{B}$ (Figure 22.10). Each charge carrier moves with the drift velocity $\mathbf{v}_d$ and feels a force $q\mathbf{v}_d \times \mathbf{B}$. The segment contains nAL carriers, where n is the number per unit volume, so the total force is

$$\mathbf{F} = (q\mathbf{v}_d \times \mathbf{B})\, nAL = (nqv_dA)\, \mathbf{L} \times \mathbf{B}, \tag{22.3.6}$$

where $\mathbf{L}$ is a vector of length L pointing along the wire in the direction of the current. From Chapter 20, $nqv_dA = I$ (equation (20.3)), so

Key Equation: Magnetic force on a current-carrying wire

$$\mathbf{F} = I\mathbf{L} \times \mathbf{B}, \qquad F = ILB\sin\theta. \tag{22.8}$$

Here θ is the angle between the wire and the field. The result does not depend on the sign of the carriers: electrons drifting one way and positive charges drifting the other both give the same force for the same current. The force on the carriers is passed to the wire itself, since the carriers cannot leave it. For a wire that is curved, or a field that is not uniform, divide the wire into short pieces $d\mathbf{L}$, each feeling a force $d\mathbf{F} = I\ d\mathbf{L} \times \mathbf{B}$, and add them up. For a closed loop of any shape in a *uniform* field, the forces add to zero; but, as the next section shows, they can still exert a torque.

Example 22.9: Levitating a wire

A horizontal copper wire of diameter $1.0\ \text{mm}$ runs east–west, perpendicular to a horizontal magnetic field. What current makes the magnetic force support the wire's weight (a) in a laboratory field of $0.10\ \text{T}$, and (b) in Earth's field, taking its horizontal component to be $2.0 \times 10^{-5}\ \text{T}$? (The density of copper is $8.96 \times 10^3\ \text{kg/m}^3$.)

Solution

The mass per unit length is $\lambda = \rho A = (8.96 \times 10^3)\pi(0.50 \times 10^{-3})^2 = 7.0 \times 10^{-3}$ kg/m. For a length L, the force must balance the weight: $ILB = \lambda Lg$, so $I = \lambda g/B$.

(a) $I = (7.0 \times 10^{-3})(9.8)/0.10 = \boxed{0.69\ \text{A}}$.

(b) $I = (7.0 \times 10^{-3})(9.8)/(2.0 \times 10^{-5}) = \boxed{3.4 \times 10^3\ \text{A}}$. Such a current would melt the thin wire almost at once (Section 20.4). Earth's field is far too weak to levitate wires, although the force it exerts on power lines is measurable. The current must flow in the right direction, too: the force $I\mathbf{L} \times \mathbf{B}$ must point up.

Checkpoint 22.3

A straight wire carries a current due east through a uniform magnetic field. What is the direction of the force on the wire if the field points (a) due north, (b) straight up, (c) due east?

Answer

Torque on a Current Loop

Although the net force on a current loop in a uniform field is zero, the forces on different sides need not act along the same line, and so they can exert a *torque*. This torque turns every electric motor.

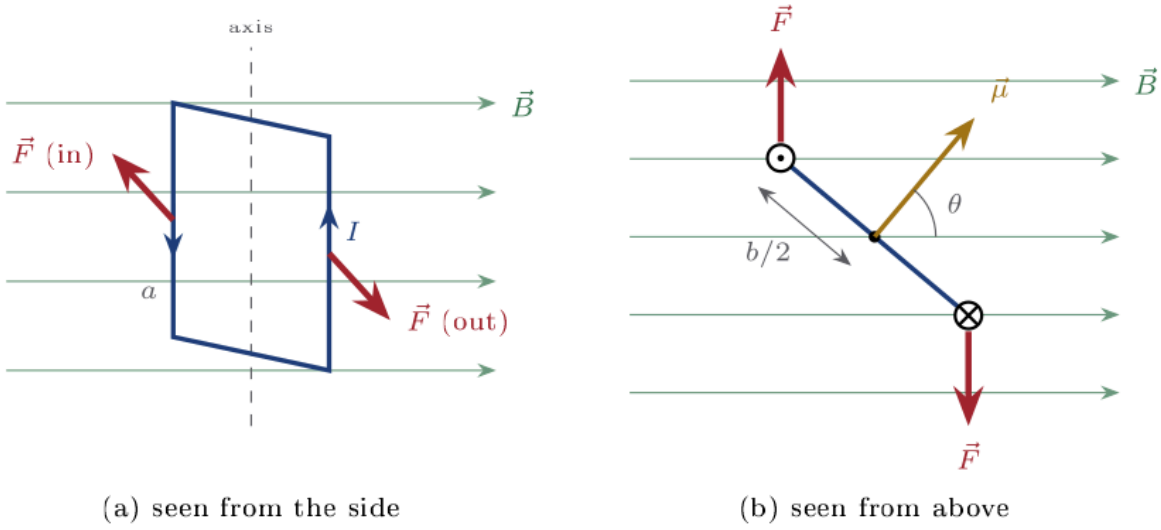


Figure 22.11. Torque on a current loop. (a) Seen from the side, the forces on the two sides of length a are equal and opposite but do not act along the same line. (b) Seen from above, they form a couple with lever arm $(b/2)\sin\theta$ each, which turns the magnetic moment μ toward the direction of $\mathbf{B}$.

•Figure description

Two views of a rectangular current loop in a uniform magnetic field B pointing to the right. Panel a, a perspective view from the side: the loop is a vertical rectangle that can rotate about a vertical axis through its center. Current I circulates around it. The two vertical sides, of length a, feel magnetic forces F of equal size in opposite directions, one out of the page and one into the page. Panel b, a view from above, looking down the axis: the loop appears edge-on as a line segment of width b through the axis. At one end of the segment the current comes out of the page, shown as a dot, and the force on it points up the page; at the other end the current goes into the page, shown as a cross, and the force on it points down. The magnetic moment mu is drawn perpendicular to the segment, making an angle theta with B. The forces form a couple that turns mu toward B.

Take a rectangular loop with sides a and b, carrying current I, free to turn about an axis through its center, in a uniform field $\mathbf{B}$ perpendicular to the axis (Figure 22.11). The two sides of length a run perpendicular to $\mathbf{B}$, so each feels a force of magnitude IaB, in opposite directions, perpendicular both to the side and to the field. The other two sides feel forces along the axis, which cancel and produce no torque. Let θ be the angle between the field and the normal to the plane of the loop. The lever arm of each of the

two forces about the axis is $(b/2)\sin\theta$, so the total torque is

$$\tau = 2\,(IaB)\,\frac{b}{2}\sin\theta = IabB\sin\theta = IAB\sin\theta, \tag{22.3.7}$$

where $A = ab$ is the area of the loop. The result holds for a flat loop of any shape. For a coil of N turns, the torque is N times larger.

The magnetic dipole moment. All the properties of the loop that matter here, its current, its area, its number of turns, and its orientation, can be collected in a single vector, the **magnetic dipole moment**:

Key Equation: Magnetic dipole moment of a current loop

$$\mu = NIA\,\hat{n}, \tag{22.9}$$

where $\hat{n}$ is the unit vector normal to the loop, with direction given by another right-hand rule: curl the fingers of your right hand around the loop in the direction of the current, and your thumb points along $\hat{n}$. The unit of μ is the $\mathrm{A}\cdot\mathrm{m}^2$. The torque and the potential energy then take exactly the forms we found for an electric dipole in an electric field (Section 16.5), with μ in place of $\mathbf{p}$ and $\mathbf{B}$ in place of $\mathbf{E}$:

Key Equation: Torque and energy of a magnetic dipole

$$\tau = \mu \times \mathbf{B}, \qquad U = -\mu \cdot \mathbf{B}. \tag{22.10}$$

The torque always tends to turn μ toward $\mathbf{B}$, and the potential energy is lowest, $-\mu B$, when μ points along the field. A compass needle is a small magnetic dipole, and it swings to line up with Earth's field for exactly this reason. The same is true of atoms, electrons, and nuclei, all of which have magnetic moments (Chapter 23); lining up the moments of hydrogen nuclei in a strong field is the first step in making an MRI image.

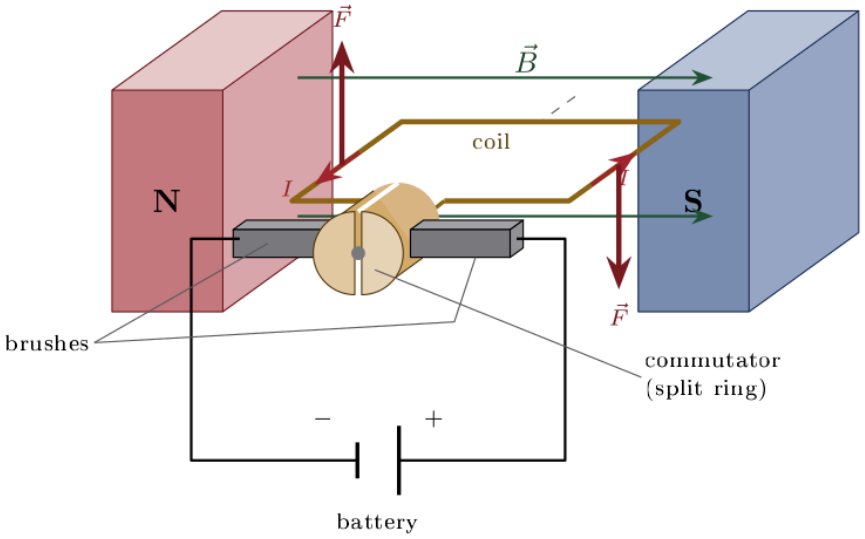


Figure 22.12. A simple DC motor. The battery drives current through fixed brushes into a split-ring commutator that turns with the coil. The forces on the two sides of the coil parallel to the axis are opposite, so they turn the coil. Every half turn the commutator halves trade brushes, reversing the current in the coil, so the torque keeps turning it the same way.

•Figure description

A simple direct-current motor in perspective. A north magnetic pole block, N, is at the left and a south pole block, S, at the right, with magnetic field lines B running from N to S across the gap. A flat rectangular coil lies horizontally in the gap and can rotate about an axis that runs from front to back through its center. On the right side of the coil the current I flows toward the back and the magnetic force F points down; on the left side the current flows toward the front and the force points up, so the coil turns. At the front, the two ends of the coil connect to the two halves of a split ring, the commutator, mounted on the axis. Two fixed brushes press on the commutator from the left and right, and wires run from the brushes down to a battery below.

Electric motors. In an **electric motor**, a coil on a rotating shaft sits in a magnetic field. The torque turns the coil toward alignment, but just as μ lines up with $\mathbf{B}$, a sliding contact called a **commutator** reverses the current in the coil. The moment now points the other way, and the torque keeps the coil turning in the same direction for another half turn, when the current reverses again. The commutator is a ring on the shaft split into two halves, one connected to each end of the coil, and two fixed brushes press on it to carry current in from the battery (Figure 22.12). Each time the coil passes alignment, the halves swap brushes, and that swap is what reverses the current. Practical motors use many coils at different angles, so that the torque is nearly constant, and in modern "brushless" motors the reversal is done electronically.

Example 22.10: The torque on a motor coil

A rectangular coil of 100 turns, $4.0\ \text{cm}$ by $6.0\ \text{cm}$, carries $2.0\ \text{A}$ in a field of $0.50\ \text{T}$. Find (a) its magnetic moment, (b) the largest torque on it, and (c) the work needed to turn it from the orientation of lowest energy to that of highest energy.

Solution

(a) $\mu = NIA = (100)(2.0)(0.040)(0.060) = \boxed{0.48\ \text{A}\cdot\text{m}^2}$.

(b) The torque is largest when μ is perpendicular to $\mathbf{B}$: $\tau_{\max} = \mu B = (0.48)(0.50) = \boxed{0.24\ \text{N}\cdot\text{m}}$.

(c) U ranges from $-\mu B$ to $+\mu B$, so the work is $2\mu B = \boxed{0.48\ \text{J}}$.

The Hall Effect

When a current flows through a conductor in a magnetic field, the magnetic force pushes the moving charge carriers toward one side of the conductor. In 1879 Edwin Hall, a graduate student at Johns Hopkins, detected the result: a small voltage across the conductor, perpendicular both to the current and to the field. The **Hall effect** gives a direct way to measure magnetic fields, and it answers a question that nothing in Chapter 20 could: what is the sign of the charge carriers?

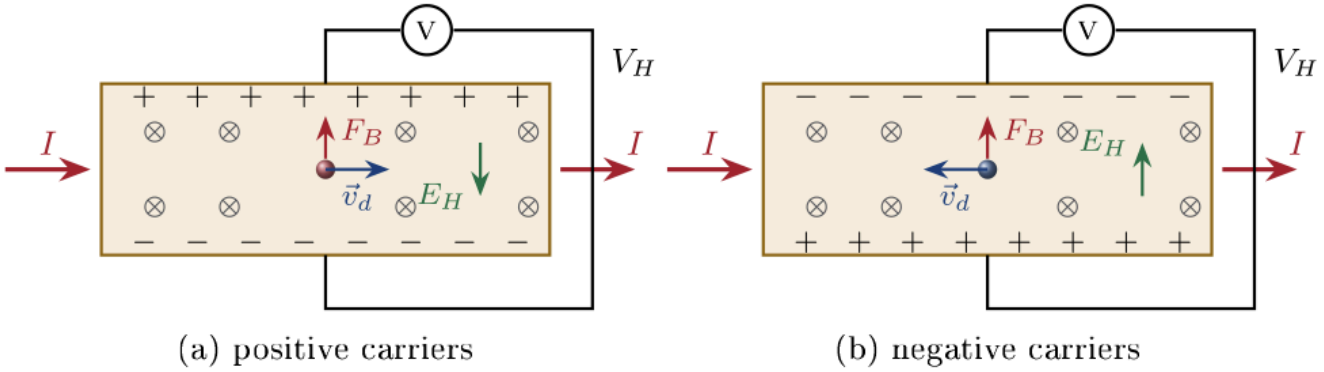


Figure 22.13. The Hall effect in a strip carrying current I to the right, with $\mathbf{B}$ into the page. The magnetic force pushes the carriers toward the top edge whatever their sign. (a) Positive carriers make the top edge positive. (b) Negative carriers make it negative. The sign of the Hall voltage reveals the sign of the carriers.

•Figure description

Two panels, each showing a flat horizontal conducting strip of width w carrying current I to the right, in a magnetic field pointing into the page, shown by crosses in circles. Panel a, positive carriers: the carriers drift to the right and the magnetic force pushes them up, so positive charge collects along the top edge and negative charge along the bottom edge; the Hall field E sub H points down. Panel b, negative carriers: electrons drift to the left, and the magnetic force pushes them up too, so negative charge collects along the top edge and positive charge along the bottom; the Hall field points up. In both panels a voltmeter connected between the top and bottom edges reads the Hall voltage V sub H, with opposite signs in the two cases.

Consider a flat strip of width w and thickness t, carrying current I along its length, in a field $\mathbf{B}$ perpendicular to its face (Figure 22.13). Suppose first that the carriers are positive, drifting in the direction of the current. The force $q\mathbf{v}_d \times \mathbf{B}$ pushes them toward one edge of the strip, which becomes positively charged, while the opposite edge is left negative. The separated charges set up a transverse electric field E_H across the strip, which pushes back on the carriers. Charge builds up until the electric force balances the magnetic force, $qE_H = qv_dB$. The potential difference across the width of the strip is then the **Hall voltage**, $V_H = E_H w = v_d B w$

. Using $I = nqv_d A$ with $A = wt$ to eliminate the drift speed,

Key Equation: Hall voltage

$$V_H = \frac{IB}{nqt}. \tag{22.11}$$

If the carriers are negative, drifting *against* the current, the magnetic force on them points toward the same edge (both q and v_d change sign), so now negative charge collects there, and the Hall voltage has the opposite sign. The sign of the Hall voltage reveals the sign of the carriers. For most metals it confirms that the carriers are electrons. But for some metals, and for p-type semiconductors, the Hall voltage has the positive sign: the current is carried by "holes," missing electrons that behave like positive charges. Once the sign is known, the size of V_H gives the carrier density n.

Example 22.11: Hall voltages in a metal and a semiconductor

(a) A copper strip $0.10\ \mathrm{mm}$ thick carries $10\ \mathrm{A}$ in a field of $1.0\ \mathrm{T}$ perpendicular to its face. Find the Hall voltage. ($n = 8.49 \times 10^{28}\ \mathrm{m}^{-3}$.) (b) Repeat for a semiconductor strip of the same thickness with $n = 1.0 \times 10^{22}\ \mathrm{m}^{-3}$, carrying only $1.0\ \mathrm{mA}$.

Solution

(a) For the copper strip,

$$V_H = \frac{IB}{nqt} = \frac{(10)(1.0)}{(8.49 \times 10^{28})(1.602 \times 10^{-19})(1.0 \times 10^{-4})} = \boxed{7.4\,\mu\mathrm{V}}, \tag{22.3.8}$$

tiny even for a large current and a strong field.

(b) For the semiconductor,

$$V_H = \frac{(1.0 \times 10^{-3})(1.0)}{(1.0 \times 10^{22})(1.602 \times 10^{-19})(1.0 \times 10^{-4})} = \boxed{6.2\,\mathrm{mV}}, \tag{22.3.9}$$

nearly a thousand times larger with ten thousand times less current. Because a semiconductor has so few carriers, they must drift much faster to carry a current, and the magnetic force on them is correspondingly larger. This is why practical Hall sensors are made of semiconductors.

Hall sensors are everywhere: they are the magnetometers that let a phone act as a compass, they sense the position of the rotor in brushless motors, they measure wheel speed for a car's antilock brakes, and they detect whether a laptop's lid is closed. Because V_H is proportional to B, a calibrated Hall probe is also the standard laboratory instrument for measuring magnetic fields.

Checkpoint 22.4

A Hall probe is held in a field and gives a reading of $+2.0\ \mathrm{mV}$. (a) The probe is turned over, so the field passes through it in the opposite direction. What does it read? (b) The probe is rotated so that its face is parallel to the field. What does it read?

Answer

Answers to Checkpoints

Checkpoint 22.3. Use $\mathbf{F} = I\mathbf{L} \times \mathbf{B}$ with $\mathbf{L}$ pointing east. Take east as $+x$, north as $+y$, and up as $+z$. (a) $\hat{\imath} \times \hat{\jmath} = \hat{k}$: the force points straight *up*. (b) $\hat{\imath} \times \hat{k} = -\hat{\jmath}$: the force points due *south*. (c) The current is parallel to the field, so the cross product is zero: there is no force. In every case the force is perpendicular to both the wire and the field.

Checkpoint 22.4. (a) $-2.0\ \mathrm{mV}$. The Hall voltage is proportional to the component of $\mathbf{B}$ perpendicular to the probe's face, and turning the probe over reverses that component, so the reading changes sign. (b) Zero. With its face parallel to the field, the field has no component perpendicular to the face, so the carriers feel no sideways force and no Hall voltage builds up. This is how a

Hall probe is used to find the direction of a field: rotate it until the reading is largest.

This page titled 22.3: Applications of Magnetic Fields is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22.S: The Nature of Magnetic Fields (Summary)

Summary

- Magnets have north and south poles, which cannot be separated. A magnetic field $\mathbf{B}$ is described by field lines that run from N to S outside a magnet and form closed loops. The SI unit is the tesla; $1\ \text{G} = 10^{-4}\ \text{T}$, and Earth's field is about $50\ \mu\text{T}$.
- A charge moving through a magnetic field feels a force $\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}$, of magnitude $|q|vB\sin\theta$, perpendicular to both $\mathbf{v}$ and $\mathbf{B}$. It does no work, so it changes a particle's direction but not its speed.
- In a uniform field, a charged particle moving perpendicular to $\mathbf{B}$ travels in a circle of radius $r = mv/|q|B$, with the **cyclotron frequency** $f = |q|B/2\pi m$, independent of speed. A velocity component along $\mathbf{B}$ makes the path a helix.
- The **Lorentz force** is $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. Crossed fields pass particles of speed $v = E/B$ (velocity selector); a mass spectrometer sorts them by $r = mv/qB$.
- Thomson measured e/m for cathode rays with crossed fields and so discovered the electron. Cyclotrons accelerate particles in a spiral at fixed frequency, reaching $K = q^2B^2R^2/2m$; synchrotrons use a fixed ring and a rising field, with $r = p/|q|B$.
- A wire carrying current I feels a force $\mathbf{F} = I\mathbf{L} \times \mathbf{B}$. A current loop has a **magnetic dipole moment** $\mu = NIA\ \hat{n}$, and feels a torque $\tau = \mu \times \mathbf{B}$, with energy $U = -\mu \cdot \mathbf{B}$. This torque drives electric motors.
- The **Hall effect**: in a field, the carriers in a current are pushed to one side, producing a Hall voltage $V_H = IB/nqt$, whose sign gives the sign of the carriers.

Key Equations

Magnetic force on a moving charge

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B}, \qquad F_B = |q|vB\sin\theta \tag{22.S.1}$$

Circular motion in a uniform field

$$r = \frac{mv}{|q|B}, \qquad f = \frac{|q|B}{2\pi m} \tag{22.S.2}$$

Lorentz force; velocity selector

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right), \qquad v = \frac{E}{B} \tag{22.S.3}$$

Cyclotron energy

$$K = \frac{q^2B^2R^2}{2m} \tag{22.S.4}$$

Force on a current-carrying wire

$$\mathbf{F} = I\mathbf{L} \times \mathbf{B} \tag{22.S.5}$$

Magnetic dipole moment, torque, and energy

$$\mu = NIA\hat{n}, \qquad \tau = \mu \times \mathbf{B}, \qquad U = -\mu \cdot \mathbf{B} \tag{22.S.6}$$

Hall voltage

$$V_H = \frac{IB}{nqt} \tag{22.S.7}$$

This page titled 22.S: The Nature of Magnetic Fields (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

22.E: The Nature of Magnetic Fields (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $e = 1.602 \times 10^{-19}$ C, $m_e = 9.11 \times 10^{-31}$ kg, $m_p = 1.673 \times 10^{-27}$ kg, and $1\ \text{u} = 1.661 \times 10^{-27}$ kg.

Section 22.1 — Magnetic Fields

22.1 ★ An alpha particle ($q = +2e$) moves at $1.5 \times 10^{7}\ \text{m/s}$ perpendicular to a $0.80\ \text{T}$ field. Find the magnetic force on it.

22.2 ★ A proton moving at $3.0 \times 10^{5}\ \text{m/s}$ feels a magnetic force of $2.4 \times 10^{-15}\ \text{N}$ when it moves at 40° to a uniform field. Find the field.

22.3 ★★ *[Conceptual.]* Each panel below shows a charged particle moving through a magnetic field. For each, give the direction of the magnetic force, or say that it is zero.



•Figure description
Four small panels. Panel a: a positive charge moving to the right in a field into the page. Panel b: a negative charge moving up the page in a field out of the page. Panel c: a positive charge moving to the right in a field that points to the right. Panel d: a negative charge moving into the page, shown as a cross in a circle at the particle, in a field that points up the page.

22.4 ★★ A particle of charge $q = -3.0\ \mu\text{C}$ has velocity $\mathbf{v} = (2.0\ \hat{\imath} - 1.0\ \hat{k}) \times 10^{4}\ \text{m/s}$ in a field $\mathbf{B} = (0.50\ \hat{\jmath})\ \text{T}$. Find the force on it, in unit-vector form.

22.5 ★★ *[Conceptual.]* An electron beam in an old television tube travels horizontally from south to north. In which direction is it deflected by Earth's magnetic field, whose horizontal component points north and whose vertical component, in the Northern Hemisphere, points down?

Section 22.2 — Motion of a Charged Particle in Magnetic Fields

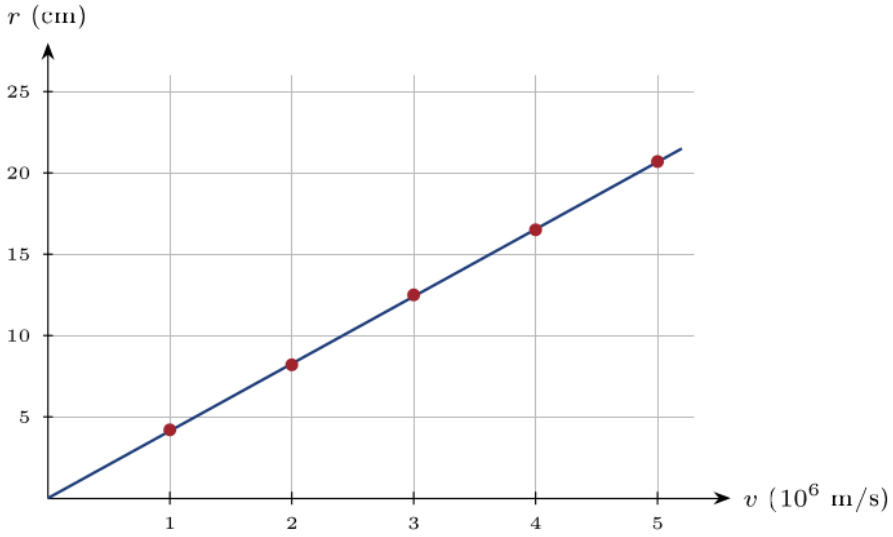
22.6 ★ A proton moves in a circle of radius $12\ \text{cm}$ perpendicular to a $0.40\ \text{T}$ field. Find its speed and its period.

22.7 ★ An electron in a uniform field circles 2.45×10^{9} times per second, the frequency of a microwave oven. What is the field? (Microwave ovens generate their waves with electrons circling in a magnetic field, in a tube called a magnetron.)

22.8 ★★ An electron is accelerated from rest through a potential difference of $500\ \text{V}$, then enters a uniform field of $2.0\ \text{mT}$ perpendicular to its velocity. Find the radius of its path.

22.9 ★★ A proton and an alpha particle ($q = 2e$, $m = 4m_p$) move through the same uniform field. Find the ratio of their orbit radii if they have (a) the same speed, (b) the same kinetic energy, (c) the same momentum.

22.10 ★★ *[Graph-reading.]* Particles of one kind are fired perpendicular to a uniform $0.25\ \text{T}$ field at different speeds, and the radius of each particle's circular path is measured. The results are plotted below.



•Figure description
A graph of orbit radius r in centimeters, from 0 to 25, against speed v in units of ten to the sixth meters per second, from 0 to 5. The data points lie on a straight line through the origin, rising to 20.7 centimeters at a speed of 5.0 times ten to the sixth meters per second.

a. Find the slope of the graph, and use it to find the charge-to-mass ratio of the particles.
b. Assuming the particles carry charge $+e$, what are they?

22.11 ★★ An electron moves at $3.0 \times 10^6\ \mathrm{m/s}$ at 60° to a uniform field of $5.0\ \mathrm{mT}$. Find the radius and the pitch of its helix.

22.12 ★★★ Show that the angular momentum, about the center of its orbit, of a particle moving in a circle perpendicular to a uniform field is $L = |q|Br^2$, and that the magnetic flux through its orbit, $\pi r^2 B$, is proportional to L. (This proportionality is behind the "adiabatic invariance" that traps particles in magnetic bottles.)

Section 22.2.1 — The Lorentz Force

22.13 ★ A velocity selector has $E = 2.4 \times 10^4\ \mathrm{V/m}$. What magnetic field selects particles of speed $6.0 \times 10^5\ \mathrm{m/s}$?

22.14 ★★ A proton moves with velocity $\mathbf{v} = (2.0 \times 10^5\ \mathrm{m/s})\ \hat{\imath}$ through the fields $\mathbf{E} = (4.0 \times 10^4\ \mathrm{V/m})\ \hat{\jmath}$ and $\mathbf{B} = (0.10\ \mathrm{T})\ \hat{k}$. Find the Lorentz force on it. What velocity along $\hat{\imath}$ would give zero force?

22.15 ★★ *[Real-world: isotope separation.]* In a mass spectrometer, singly charged ions of uranium-235 and uranium-238 ($m = 235.0\ \mathrm{u}$ and $238.1\ \mathrm{u}$) leave a velocity selector at $3.0 \times 10^5\ \mathrm{m/s}$ and enter a field of $0.60\ \mathrm{T}$. How far apart do they land? Why was this method so slow for producing bomb-grade uranium in the 1940s?

22.16 ★★★ A particle of charge q and mass m starts from rest at the origin in uniform crossed fields $\mathbf{E} = E\ \hat{\jmath}$ and $\mathbf{B} = B\ \hat{k}$. Describe its motion qualitatively, and show that its average velocity is E/B along $\hat{\imath}$, perpendicular to both fields. (Its path is a cycloid; this "$E \times B$ drift" is important in plasmas.)

Section 22.3.1 — Discovery of the Electron

22.17 ★★ In a Thomson experiment, electrons accelerated through $1.0\ \mathrm{kV}$ pass undeflected through plates $1.5\ \mathrm{cm}$ apart with $240\ \mathrm{V}$ between them, in a magnetic field B. Find the electrons' speed and the required B.

22.18 ★★ *[Conceptual.]* Thomson found the same q/m for cathode rays whatever the cathode metal and whatever gas remained in the tube. Why was this so important for the conclusion that electrons are a constituent of all atoms?

Section 22.3.2 — Cyclotrons and Synchrotrons

22.19 ★ What frequency is needed to accelerate deuterons ($q = e$, $m = 3.34 \times 10^{-27}\ \text{kg}$) in a cyclotron with a $1.2\ \text{T}$ field?

22.20 ★★ *[Real-world: medicine.]* A medical cyclotron accelerates protons to $18\ \text{MeV}$ to make the fluorine-18 used in PET scans. If its field is $1.9\ \text{T}$, what radius must its dees have?

22.21 ★★ Show that doubling the field of a cyclotron quadruples the final kinetic energy of its particles, and doubling the radius does the same. Which is easier in practice?

22.22 ★★★ *[Real-world: particle physics.]* Electrons in a synchrotron light source have an energy of $3.0\ \text{GeV}$, so $p \approx E/c$. The ring's bending magnets have a field of $1.3\ \text{T}$. Find the radius of curvature in the magnets. Then use the relativistic relation $E = \gamma mc^2$ to find γ, and explain why these electrons radiate so strongly.

Section 22.3.3 — Magnetic Force on a Current-Carrying Wire

22.23 ★ A $0.50\ \text{m}$ wire carrying $8.0\ \text{A}$ lies at 60° to a $0.30\ \text{T}$ field. Find the force on it.

22.24 ★★ *[Real-world: power lines.]* A power line carries $1000\ \text{A}$ east, where Earth's field is $50\ \mu\text{T}$ directed north and 60° below the horizontal. Find the magnitude and direction of the magnetic force on a $100\ \text{m}$ span.

22.25 ★★ A semicircular loop of wire of radius R, closed by its diameter, carries current I in a uniform field $\mathbf{B}$ perpendicular to its plane. Show that the force on the curved part is $2IRB$, equal and opposite to the force on the straight diameter.

22.26 ★★★ *[Real-world: rail guns.]* A metal bar of mass $0.20\ \text{kg}$ slides without friction on two long parallel rails $0.10\ \text{m}$ apart, carrying a current of $2.0 \times 10^4\ \text{A}$ across between them, in a field of $5.0\ \text{T}$ perpendicular to the plane of the rails. Find its acceleration, and its speed after sliding $4.0\ \text{m}$ from rest.

Section 22.3.4 — Torque on a Current Loop

22.27 ★ A circular coil of 50 turns and radius $3.0\ \text{cm}$ carries $0.80\ \text{A}$. Find its magnetic moment, and the maximum torque on it in a $0.25\ \text{T}$ field.

22.28 ★★ The coil of Example 22.10 makes an angle of 30° between μ and $\mathbf{B}$. Find the torque and the potential energy.

22.29 ★★ *[Conceptual.]* A current loop in a uniform field is in equilibrium when μ is parallel to $\mathbf{B}$ and when it is antiparallel. Which equilibrium is stable? Compare with an electric dipole in an electric field.

22.30 ★★★ A current loop of moment μ and moment of inertia I about its axis sits in a uniform field B. Show that, when turned slightly from its stable orientation and released, it oscillates with period $T = 2\pi\sqrt{I/\mu B}$. (Compare Problem 16.29. This is how a compass needle oscillates before it settles.)

Section 22.3.5 — The Hall Effect

22.31 ★ A Hall probe of thickness $0.20\ \text{mm}$ with $n = 2.0 \times 10^{22}\ \text{m}^{-3}$ carries $5.0\ \text{mA}$. What Hall voltage does it produce in a field of $0.10\ \text{T}$?

22.32 ★★ A strip of an unknown metal $0.050\ \text{mm}$ thick carries $8.0\ \text{A}$ in a field of $1.2\ \text{T}$, and shows a Hall voltage of $12\ \mu\text{V}$, with the sign expected for negative carriers. Find the density of carriers, and the number per atom if the metal has 5.9×10^{28} atoms per cubic meter.

22.33 ★★★ Show that the drift speed of the carriers in a Hall experiment is $v_d = V_H/Bw$, independent of the carrier density. Find v_d for the copper strip of Example 22.11 if it is $1.0\ \text{cm}$ wide, and compare with the drift speed you would calculate from $I = nev_dA$.

This page titled 22.E: The Nature of Magnetic Fields (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23: MAGNETIC FIELDS AND CURRENTS

On a spring morning in 1820, during a lecture in Copenhagen, Hans Christian Oersted set a compass beside a wire and closed a circuit. The needle swung sharply, to point across the wire. For decades physicists had suspected that electricity and magnetism were connected; here at last was proof. Within weeks the news reached Paris, and within months André-Marie Ampère had worked out how currents exert forces on each other, while Jean-Baptiste Biot and Félix Savart had measured the field of a current and found the law that describes it.

Chapter 22 took magnetic fields as given and asked what they do to moving charges. This chapter asks where they come from. The answer is simple and surprising: *every* magnetic field is produced by moving charge, whether the current in a wire or the motion and spin of the electrons inside a magnet. We will calculate the fields of wires, loops, and coils, first by adding up the contributions of small pieces of current, then, where the symmetry allows, with the far more efficient Ampère's law. We will find why magnetic field lines never end, and close by looking inside matter to see how atoms become magnets, and why iron is special.

Learning Objectives

- Use the Biot–Savart law to find the magnetic fields of straight wires, arcs, and loops.
- Calculate the force between two parallel currents, and explain how it defined the ampere.
- Describe the field of a magnetic dipole, and relate a current loop, a bar magnet, and the Earth.
- Apply Ampère's law to symmetric current distributions, including solenoids and toroids.
- State Gauss's law for magnetism and explain what it says about magnetic monopoles.
- Explain the magnetism of atoms and distinguish diamagnetic, paramagnetic, and ferromagnetic materials.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

23.1: Magnetic Field Due to a Current
23.2: Force Between Two Parallel Currents
23.3: Magnetic Dipoles
23.4: Ampère's Law
23.5: Solenoids and Toroids
23.6: Gauss's Law in Magnetism
23.7: Magnetic Properties of Matter
23.S: Magnetic Fields and Currents (Summary)
23.E: Magnetic Fields and Currents (Problems)

This page titled 23: Magnetic Fields and Currents is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.1: Magnetic Field Due to a Current

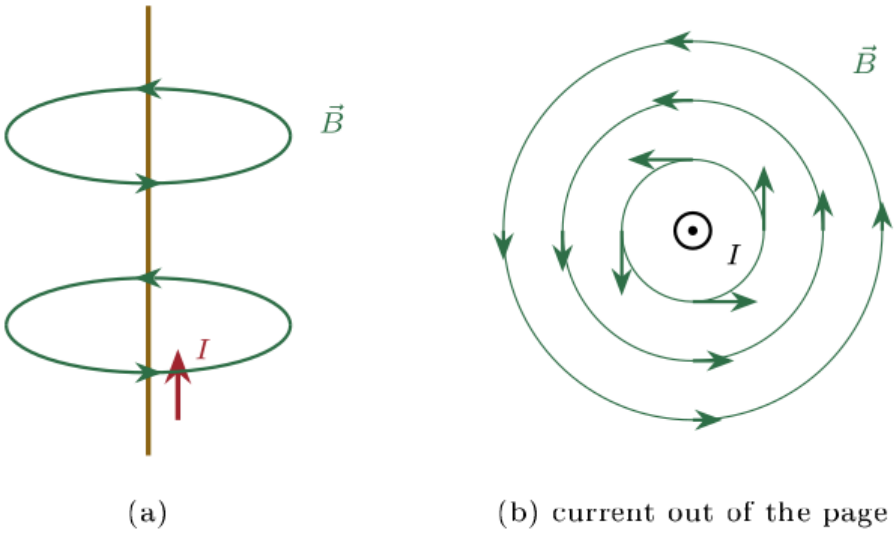


Figure 23.1. The magnetic field of a long straight wire. (a) The field lines are circles around the wire, their direction given by the right-hand rule. (b) Seen along the wire, with the current out of the page, the field circles counterclockwise and weakens with distance.

•Figure description

Two views of the magnetic field of a long straight wire carrying current I. Panel a, a perspective view: a vertical wire carries current I upward. Two circular field lines, drawn as ellipses in perspective, loop around the wire at different heights; arrows show that the field circles counterclockwise when seen from above. Panel b, a view looking along the wire, with the current coming out of the page, shown as a dot in a circle: three concentric circular field lines surround the wire, with arrows showing counterclockwise circulation, and field vectors B tangent to the circles, shorter on the outer circles.

Oersted's compass showed that a current in a long straight wire produces a magnetic field that circles the wire (Figure 23.1). The field lines are circles centered on the wire, in planes perpendicular to it. Their direction is given by a **right-hand rule**: grasp the wire with your right hand, thumb pointing along the current, and your fingers curl in the direction of $\mathbf{B}$. The field is strongest near the wire and weakens with distance.

The Biot–Savart law. To calculate the field, we proceed as we did for continuous charge distributions in Chapter 16: divide the current into small pieces, find the field of each, and add. A short length $d\mathbf{s}$ of wire carrying current I, pointing in the direction of the current, is called a *current element*. Biot and Savart found experimentally that its contribution to the field at a point a distance r away, in the direction of the unit vector $\hat{r}$ from the element to the point, is

Key Equation: The Biot–Savart law

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I\,d\mathbf{s}\times\hat{r}}{r^2}, \qquad dB = \frac{\mu_0}{4\pi}\frac{I\,ds\,\sin\theta}{r^2}. \tag{23.1}$$

This is the **Biot–Savart law**. Here θ is the angle between $d\mathbf{s}$ and $\hat{r}$, and the constant μ_0 is the **permeability of free space**,

$$\mu_0 = 4\pi\times10^{-7}\,\mathrm{T\cdot m/A} = 1.257\times10^{-6}\,\mathrm{T\cdot m/A}, \tag{23.1.1}$$

the magnetic counterpart of ϵ_0. (Before 2019, this value was exact by definition. It is now a measured quantity, equal to $4\pi\times10^{-7}$ to better than one part in a billion.) The law resembles Coulomb's law: the field falls off as $1/r^2$, and it is proportional to the strength of the source, $I\ ds$ in place of q. But the cross product makes it quite different in direction. The field of a current element is perpendicular both to the element and to $\hat{r}$, and it is zero along the line of the current itself.

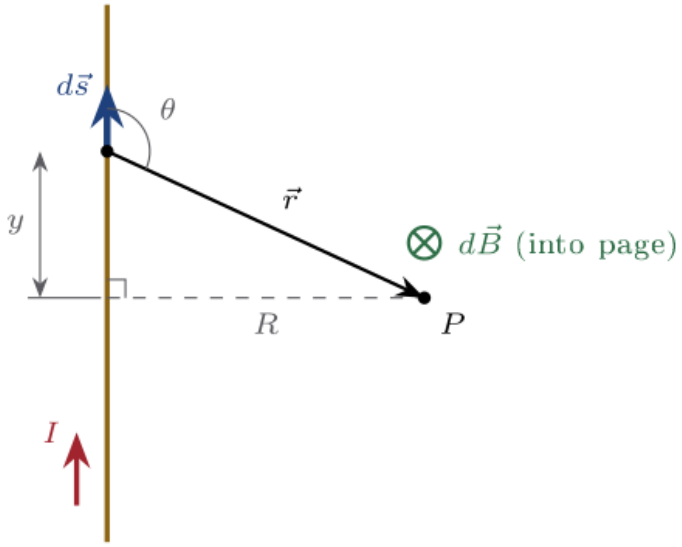


Figure 23.2. Finding the field of a long straight wire with the Biot–Savart law. Every current element produces a field at P pointing into the page, in the direction of $d\mathrm{s} \times \hat{r}$.

•Figure description

Geometry for the Biot-Savart law. A long straight wire lies along a vertical y axis and carries current I upward. A short current element d s at height y above the origin points up along the wire. A point P lies a perpendicular distance R to the right of the origin. The vector r runs from the element down and to the right to P, making an angle theta with d s. At P, the field d B of the element points into the page, shown as a cross in a circle.

A long straight wire. Let a long straight wire lie along the y axis, carrying current I in the $+y$ direction, and find the field at a point P a perpendicular distance R from it (Figure 23.2). Every element $d\mathrm{s} = dy\ \hat{\jmath}$ contributes a field pointing in the same direction, into the page at P, so we need only add magnitudes. For the element at height y, $r = \sqrt{y^2 + R^2}$ and $\sin\theta = R/r$, so

$$B = \frac{\mu_0 I}{4\pi}\int_{-\infty}^{\infty}\frac{R\,dy}{(y^2+R^2)^{3/2}} = \frac{\mu_0 I}{4\pi}\left[\frac{y}{R\sqrt{y^2+R^2}}\right]_{-\infty}^{\infty} = \frac{\mu_0 I}{4\pi}\cdot\frac{2}{R}. \tag{23.1.2}$$

The field of a long straight wire is

Key Equation: Field of a long straight wire

$$B = \frac{\mu_0 I}{2\pi R}. \tag{23.2}$$

It falls off as $1/R$, just as the electric field of a long line of charge does (equation (16.10)), and for the same geometric reason. For a wire of finite length, the same integral taken between the ends gives a smaller field (Problem 23.5).

Example 23.1: A wire and a compass

A straight wire carries a current of $10\ \mathrm{A}$. Find the magnetic field $2.0\ \mathrm{cm}$ from it, and compare it with Earth's field of about $50\ \mu\mathrm{T}$.

Solution

$$B = \frac{\mu_0 I}{2\pi R} = \frac{(4\pi\times10^{-7})(10)}{2\pi(0.020)} = \boxed{1.0\times10^{-4}\ \mathrm{T}}, \tag{23.1.3}$$

twice Earth's field. A compass held near a wire carrying a few amperes points noticeably away from north, which is exactly what Oersted saw. It is also why a compass near a car's wiring, or an electronic device, gives misleading readings. (Wires carrying equal currents in opposite directions, like the two conductors of a household cable, produce much smaller fields, since their fields nearly cancel.)

Arcs and loops. For a wire bent into a circular arc of radius R, the field at the center of the arc is easy to find. Every element is the same distance R from the center, and every $d\mathbf{s}$ is perpendicular to $\hat{r}$, so each contributes $dB = (\mu_0/4\pi)I\ ds/R^2$, all in the same direction, perpendicular to the plane of the arc. For an arc subtending an angle ϕ (in radians), the total length is $R\phi$, and

$$B = \frac{\mu_0 I \phi}{4\pi R}. \tag{23.3}$$

For a complete circular loop, $\phi = 2\pi$, and the field at its center is

Key Equation: Field at the center of a circular loop

$$B = \frac{\mu_0 I}{2R} \qquad (N \text{ turns: } B = \frac{\mu_0 NI}{2R}). \tag{23.4}$$

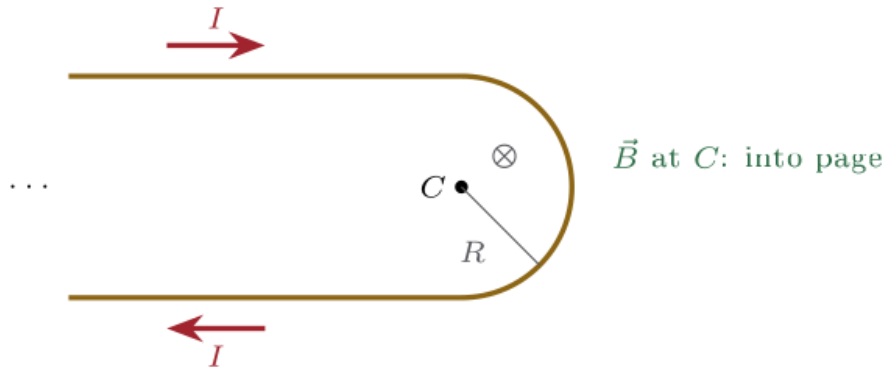


Figure 23.3. A hairpin. The two straight sections and the semicircle all produce fields into the page at the center C.

•Figure description

A hairpin-shaped wire. Two long horizontal straight wires extend from the left, one above the other, separated by twice the radius R. At their right ends they are joined by a semicircle bulging to the right. Current I flows to the right along the top wire, around the semicircle, and back to the left along the bottom wire. The center C of the semicircle lies midway between the straight wires. At C, the field B points into the page.

Its direction follows from another form of the right-hand rule: curl the fingers of your right hand around the loop in the direction of the current, and your thumb points along $\mathbf{B}$ at the center. This is the same rule that gave the direction of the loop's magnetic moment μ in Section 22.3.

Example 23.2: A hairpin

A long wire is bent into the hairpin shape of Figure 23.3: two long parallel straight sections joined by a semicircle of radius $R = 2.0\ \text{cm}$. It carries $5.0\ \text{A}$. Find the field at the center C of the semicircle.

Solution

Each straight section is a "semi-infinite" wire, ending at the point directly beside C. By symmetry it produces half the field of an infinite wire, $\mu_0 I/4\pi R$. (The integral of the long-wire calculation, taken from 0 to ∞ instead of from $-\infty$ to ∞, gives half the result.) The semicircle gives equation (23.3) with $\phi = \pi$, that is, $\mu_0 I/4R$. By the right-hand rule, all three contributions point into the page. So

$$B = 2\,\frac{\mu_0 I}{4\pi R} + \frac{\mu_0 I}{4R} = \frac{\mu_0 I}{4\pi R}(2+\pi) = \frac{(4\pi \times 10^{-7})(5.0)}{4\pi(0.020)}(5.14) = \boxed{1.3 \times 10^{-4}\ \text{T}}, \tag{23.1.4}$$

into the page.

Checkpoint 23.1

A long straight wire lies in the plane of the page, carrying current to the right. What is the direction of its magnetic field (a) at a point just above it in the page, (b) at a point just below it, and (c) at a point on the wire's own line, far to its right?

Answer

Answers to Checkpoints

Checkpoint 23.1. By the right-hand rule, point your thumb along the current (to the right) and your fingers curl around the wire. (a) Just above the wire, in the page, the field points *out of* the page. (b) Just below, it points *into* the page. (c) Zero. For any point on the wire's own line, each current element $d\mathbf{s}$ is parallel to $\hat{r}$, so $d\mathbf{s} \times \hat{r} = 0$ in the Biot–Savart law: a current produces no field straight ahead of itself.

This page titled 23.1: Magnetic Field Due to a Current is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.2: Force Between Two Parallel Currents

A current produces a magnetic field, and a magnetic field exerts a force on a current. So two currents exert forces on each other. Ampère discovered this within weeks of hearing of Oersted's experiment.

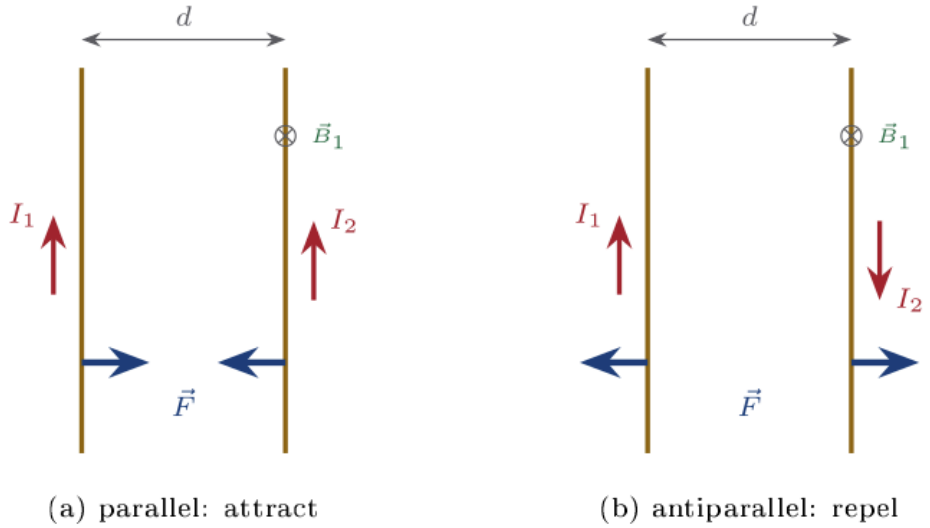


Figure 23.4. The force between parallel currents. (a) The field of wire 1 at wire 2 points into the page, and the force $I_2\mathbf{L} \times \mathbf{B}_1$ on wire 2 points toward wire 1: parallel currents attract. (b) Reversing one current reverses the forces: antiparallel currents repel.

•Figure description

Two panels, each showing two long vertical parallel wires a distance d apart. Panel a: both currents flow upward. The field of wire 1 at wire 2 points into the page. The force on wire 2 points left, toward wire 1, and the force on wire 1 points right, toward wire 2: the wires attract. Panel b: the current in wire 1 flows up and the current in wire 2 flows down. The forces point away from each other: the wires repel.

Let two long parallel wires a distance d apart carry currents I_1 and I_2 in the same direction (Figure 23.4a). Wire 1 produces a field $B_1 = \mu_0 I_1/2\pi d$ at the position of wire 2, perpendicular to wire 2. By equation (22.8), a length L of wire 2 feels a force $F = I_2 L B_1$, and the right-hand rules show that it points toward wire 1. The force per unit length is

Key Equation: Force between parallel currents

$$\frac{F}{L} = \frac{\mu_0 I_1 I_2}{2\pi d}. \tag{23.5}$$

By Newton's third law, wire 1 feels an equal and opposite force. The rule is the reverse of the rule for charges: *parallel currents attract, and antiparallel currents repel* (Figure 23.4b).

Defining the ampere. From 1948 to 2019, this force defined the unit of current. One ampere was the current that, flowing in each of two long parallel wires one meter apart in vacuum, produces a force of exactly 2×10^{-7} newton per meter of length. That is why μ_0 was exactly $4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$: the definition built it in. Since 2019 the ampere has instead been defined by fixing the value of the elementary charge e, so that an ampere is a definite number of elementary charges per second, and μ_0 became a measured constant. The measured value is the same to nine decimal places.

Example 23.3: A short circuit in a substation

Two parallel copper busbars in an electrical substation are $10\ \mathrm{cm}$ apart and carry currents in opposite directions. Find the force per meter between them (a) at the normal current of $2000\ \mathrm{A}$, and (b) during a short circuit, when the current briefly rises to $40\ \mathrm{kA}$.

Solution

(a) The force per unit length is

$$\frac{F}{L} = \frac{\mu_0 I^2}{2\pi d} = \frac{(4\pi \times 10^{-7})(2000)^2}{2\pi(0.10)} = \boxed{8.0\,\mathrm{N/m}}, \tag{23.2.1}$$

pushing the bars apart.

(b) The force grows as I^2, so a current twenty times larger gives four hundred times the force: $\boxed{3.2 \times 10^3\,\mathrm{N/m}}$. That is the weight of more than $300\ \mathrm{kg}$ on every meter of bar. Busbars and their supports are designed to survive these brief but violent magnetic forces, and a short circuit is often heard as a loud bang as the conductors jerk apart.

This page titled 23.2: Force Between Two Parallel Currents is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.3: Magnetic Dipoles

A current loop behaves like a small magnet: in a magnetic field it feels a torque that turns its moment μ toward the field (Section 22.3). It also *produces* a field like that of a magnet. To see this, find the field of a circular loop at points on its axis.

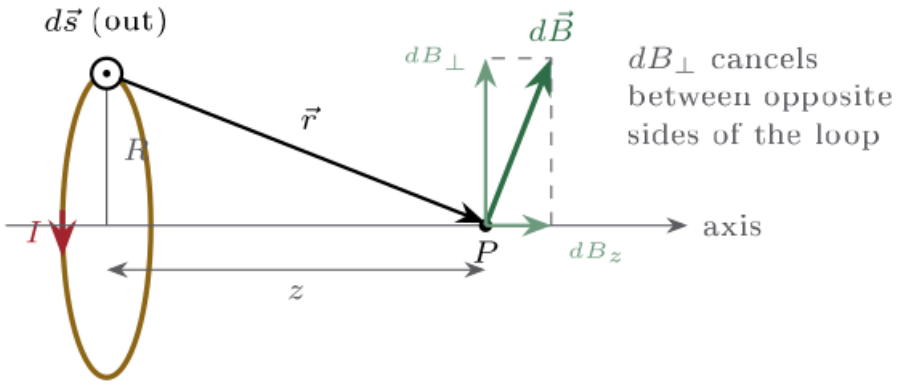


Figure 23.5. The field on the axis of a current loop. Each element contributes a field perpendicular to r; the components perpendicular to the axis cancel in pairs, leaving a field along the axis.

•Figure description

A circular current loop of radius R drawn in perspective as a narrow vertical ellipse, with its axis running horizontally to the right. At the top of the loop, a current element d s points out of the page, shown as a dot in a circle. A point P lies on the axis a distance z to the right of the loop's center. The vector r runs from the top element to P. At P, the element's field d B is perpendicular to r, pointing up and to the right; it is resolved into a component along the axis and a component perpendicular to the axis. A note says that the perpendicular components from opposite sides of the loop cancel.

The field on the axis of a loop. Take a loop of radius R carrying current I, and a point P on its axis a distance z from its center (Figure 23.5). Every element $d\mathrm{s}$ is perpendicular to the vector r from it to P, and at the same distance $r = \sqrt{z^2 + R^2}$, so each contributes a field of magnitude $dB = (\mu_0/4\pi)I\ ds/r^2$, perpendicular to r. Each $d\mathbf{B}$ has a component along the axis and a component perpendicular to it. The perpendicular components of elements on opposite sides of the loop cancel, exactly as for the electric field on the axis of a ring of charge (Section 16.3). The axial components, each $dB\cos\alpha = dB\ R/r$, all add:

$$B = \frac{\mu_0 I}{4\pi r^2}\,\frac{R}{r}\oint ds = \frac{\mu_0 I}{4\pi}\,\frac{R\,(2\pi R)}{(z^2+R^2)^{3/2}}, \tag{23.3.1}$$

so

Key Equation: Field on the axis of a current loop

$$B = \frac{\mu_0 I R^2}{2\,(z^2+R^2)^{3/2}}. \tag{23.6}$$

At the center, $z = 0$, this reduces to $\mu_0 I/2R$, equation (23.4).

Far from the loop. For $z \gg R$, the denominator becomes $2z^3$, and with the loop's magnetic moment $\mu = I\pi R^2$,

$$B \approx \frac{\mu_0}{4\pi}\,\frac{2\mu}{z^3}. \tag{23.7}$$

Compare the electric field on the axis of an electric dipole far away, $E = 2k_e p/z^3 = (1/4\pi\epsilon_0)(2p/z^3)$, equation (16.7). The two have exactly the same form, with $\mu_0/4\pi$ in place of $1/4\pi\epsilon_0$ and μ in place of p. The resemblance is not limited to the axis: far from the loop, at every point, its magnetic field has the same shape as the electric field of an electric dipole. On the perpendicular bisector, for example, it is $B = (\mu_0/4\pi)(\mu/r^3)$, pointing opposite to μ. A current loop is a **magnetic dipole**.

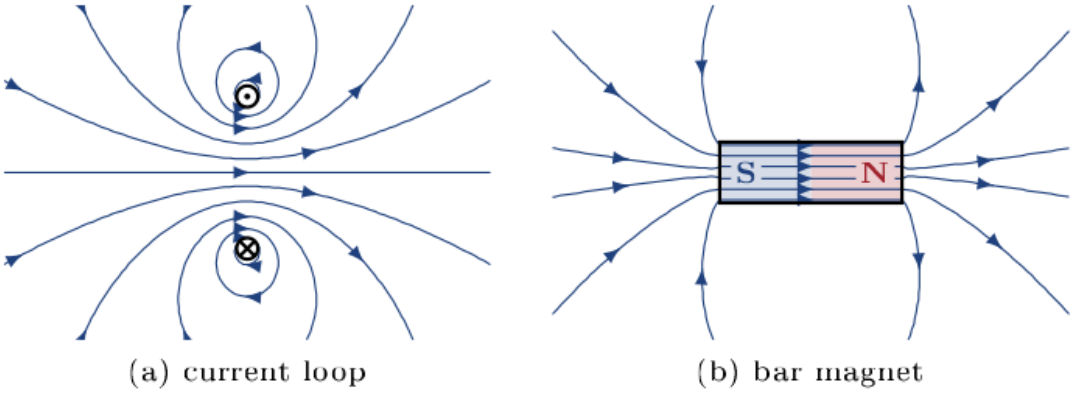


Figure 23.6. Computed magnetic field lines of (a) a current loop, seen edge-on, and (b) a bar magnet. Away from the wire and the magnet, both are dipole fields of the same shape. Both have their magnetic moments pointing to the right.

•Figure description

Two computed field-line diagrams side by side. Left: a circular current loop seen edge-on, so that the wire crosses the page at two points, one above the axis with current out of the page and one below with current into the page. Field lines pass through the loop from left to right, spread out beyond it, and curve back around the outside; close to each wire the lines form small closed circles around it. Right: a bar magnet with its south pole on the left and north pole on the right; its field lines leave the north pole, curve around, and enter the south pole. Away from the sources, the two patterns look the same.

Figure 23.6 shows computed field lines of a current loop beside those of a bar magnet. Near the wire they differ: the lines circle the wire itself. But farther away the two patterns are the same. A bar magnet is a magnetic dipole, with its moment pointing from its south pole to its north pole. So, to a good approximation, is the Earth, whose dipole moment is about $8.0 \times 10^{22}\ \mathrm{A \cdot m^2}$, tilted about 10° from its axis of rotation. The source of Earth's field is a current: electric currents in its molten iron outer core, sustained by the slow churning of the liquid metal, a self-sustaining "geodynamo."

Example 23.4: Earth's dipole field

Treat Earth as a magnetic dipole of moment $8.0 \times 10^{22}\ \mathrm{A \cdot m^2}$ at its center. Find the field at the surface ($R_E = 6.37 \times 10^6\ \mathrm{m}$) (a) at a magnetic pole, on the dipole's axis, and (b) at the magnetic equator.

Solution

(a) On the axis, from equation (23.7),

$$B = \frac{\mu_0}{4\pi}\frac{2\mu}{R_E^3} = (10^{-7})\frac{2(8.0 \times 10^{22})}{(6.37 \times 10^6)^3} = \boxed{62\,\mu\mathrm{T}}. \tag{23.3.2}$$

(b) On the equator, which lies on the dipole's perpendicular bisector, the field is half as large, $\boxed{31\,\mu\mathrm{T}}$, and points horizontally. Both agree well with measurements: Earth's field is about $60\ \mu\mathrm{T}$ near the magnetic poles and about $30\ \mu\mathrm{T}$ near the equator.

Checkpoint 23.2

A small bar magnet is replaced by a current loop with the same magnetic moment. (a) Far away, can you tell the difference by measuring the field? (b) How does the field on the axis change if you go twice as far from the loop?

Answer

Answers to Checkpoints

Checkpoint 23.2. (a) No. Far from either one, the field depends only on the magnetic moment μ and has the same dipole pattern. A bar magnet and a current loop with equal moments are indistinguishable at a distance. (Close up they differ: inside the loop the field lines run straight through, while the magnet's are hidden inside the iron.) (b) Far along the axis the dipole field falls as $1/z^3$,

so doubling the distance cuts it to $1/8$.

This page titled 23.3: Magnetic Dipoles is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.4: Ampère's Law

The Biot–Savart law gives the field of any current, but the integrals can be quite challenging. Just as Gauss's law made electric fields easy to find when the charges have enough symmetry (Chapter 17), a second law makes magnetic fields easy to find when the currents are symmetric. It concerns the *circulation* of $\mathbf{B}$ around a closed path.

Circulation around a wire. Take a circle of radius r around a long straight wire. The field of the wire is tangent to the circle everywhere, with constant magnitude $\mu_0 I/2\pi r$. Divide the circle into short steps $d\mathbf{s}$ and add up $\mathbf{B}\cdot d\mathbf{s}$ all the way around:

$$\oint \mathbf{B}\cdot d\mathbf{s} = B\,(2\pi r) = \frac{\mu_0 I}{2\pi r}(2\pi r) = \mu_0 I. \tag{23.4.1}$$

The radius has dropped out. It turns out that the result depends only on the current passing through the path, not on the shape of the path at all, and that currents outside the path contribute nothing. This is **Ampère's law**:

Key Equation: Ampère's law

$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 I_{\text{enc}}. \tag{23.8}$$

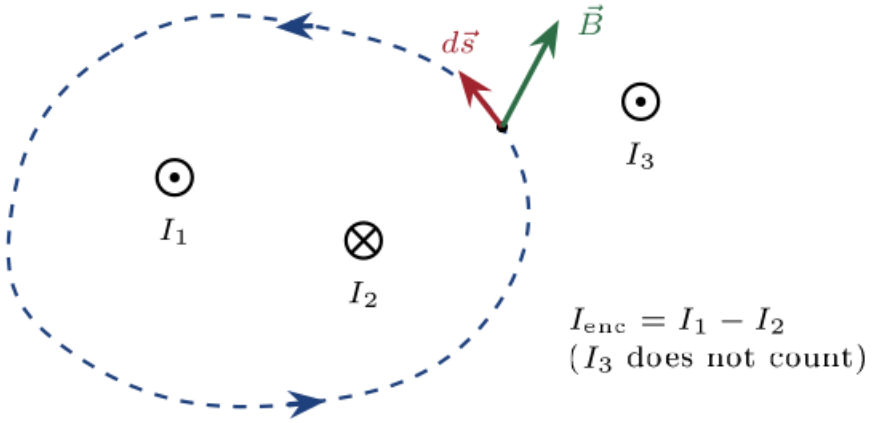


Figure 23.7. Ampère's law. For a loop traversed counterclockwise, currents out of the page count as positive. Only the currents that pass through the loop contribute to I_{enc}, although all the currents contribute to the field $\mathbf{B}$ along it.

•Figure description

An Ampèrian loop, an irregular closed curve drawn in the plane of the page and traversed counterclockwise, as shown by arrows. Three long wires pierce the page. Inside the loop, wire 1 carries current I sub 1 out of the page and wire 2 carries current I sub 2 into the page. Outside the loop, wire 3 carries current I sub 3 out of the page. At one point on the loop, a short step d s is tangent to the loop and the local field B points at an angle to it. A label reads I enclosed equals I sub 1 minus I sub 2.

The integral is taken around any closed path, called an **Ampèrian loop**, and I_{enc} is the net current passing through any surface bounded by that path. Signs follow the right-hand rule: curl the fingers of your right hand around the loop in the direction you traverse it, and currents in the direction of your thumb count as positive (Figure 23.7).

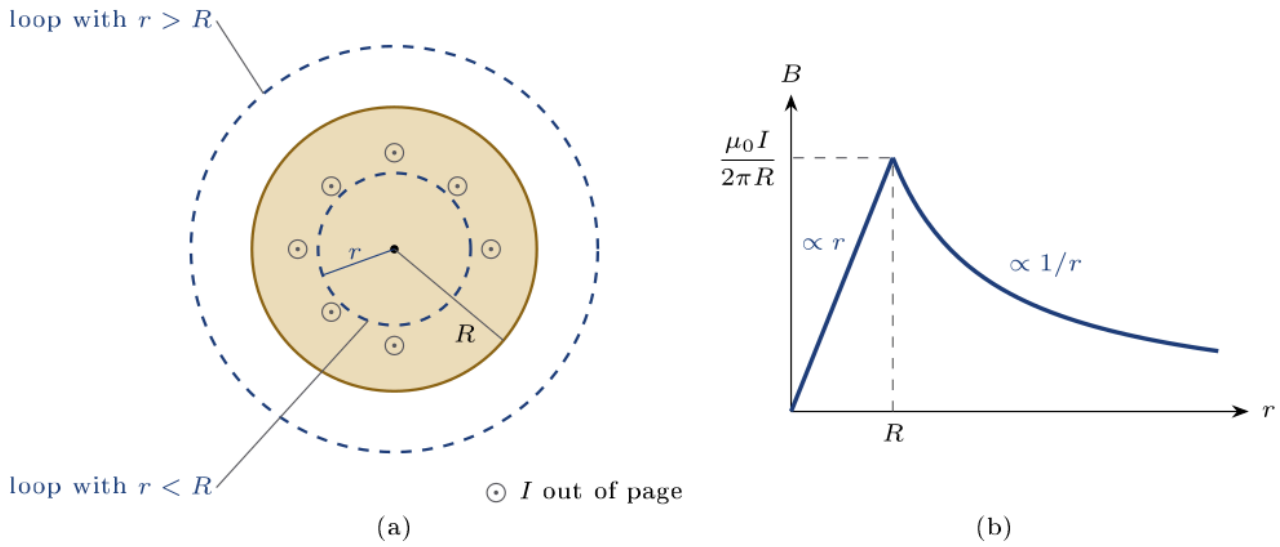


Figure 23.8. The field of a long straight wire of radius R carrying a uniform current. (a) Ampèrian loops inside and outside the wire. (b) The field grows in proportion to r inside the wire and falls off as $1/r$ outside, where it is the same as if all the current flowed along the axis.

•Figure description

Panel a: the cross-section of a long straight wire of radius R, shaded uniformly to show that the current, out of the page, is spread evenly over it. Two dashed Ampèrian circles are drawn and labeled: a loop with r less than R inside the wire, and a loop with r greater than R outside it. The radius R of the wire and the radius r of the inner loop are marked. Panel b: a graph of the magnetic field B against distance r from the axis. B rises linearly from zero at the axis to a maximum of mu zero I over two pi R at the surface r equals R, then falls off as one over r outside the wire.

Ampère's law is true for any steady current and any loop, but, like Gauss's law, it is *useful* for finding $\mathbf{B}$ only when the symmetry lets us choose a loop along which $\mathbf{B}$ is either tangent with constant magnitude, or perpendicular, so that $\mathbf{B}$ can be taken outside the integral. The main cases are the long straight wire, the solenoid, and the toroid. In its current form, Ampère's law is begging for a modification. It will make its return in Chapter 27, where its complete form tells us something profound about the underlying nature of our universe.

Inside a wire. Let a long straight wire of radius R carry current I spread uniformly over its cross-section (Figure 23.8). By symmetry, the field lines are circles centered on the axis. Outside the wire, a circular Ampèrian loop encloses the whole current, and we recover $B = \mu_0 I/2\pi r$. Inside, a circle of radius $r < R$ encloses only the fraction $\pi r^2/\pi R^2$ of the current, so

$$B(2\pi r) = \mu_0 I \frac{r^2}{R^2} \qquad \Longrightarrow \qquad B = \frac{\mu_0 I r}{2\pi R^2} \quad (r \le R). \tag{23.4.2}$$

The field grows linearly from zero at the axis to its maximum at the surface, then falls off as $1/r$ outside.

Example 23.5: Inside and outside a wire

A copper wire of radius $1.0\ \text{mm}$ carries a uniform current of $10\ \text{A}$. Find the field (a) at its surface, (b) halfway to its axis, and (c) $5.0\ \text{mm}$ from its axis.

Solution

(a) At the surface, an Ampèrian loop of radius R just encloses the whole current, so the inside and outside formulas agree: $B = \mu_0 I/2\pi R$. It helps to note that $\mu_0/2\pi = 2 \times 10^{-7}\ \text{T}\cdot\text{m/A}$. Then

$$B = \frac{\mu_0 I}{2\pi R} = \frac{(2\times 10^{-7}\,\text{T}\cdot\text{m/A})(10\,\text{A})}{1.0\times 10^{-3}\,\text{m}} = \boxed{2.0\,\text{mT}}. \tag{23.4.3}$$

(b) Halfway to the axis, $r = R/2 = 0.50\ \text{mm}$. A loop of this radius encloses only the fraction $(r/R)^2 = 1/4$ of the current, $2.5\ \text{A}$, but it is also only half as long as the loop at the surface. Using the inside result $B = \mu_0 I r/2\pi R^2$,

$$B = \frac{(2 \times 10^{-7}\,\mathrm{T \cdot m/A})(10\,\mathrm{A})(0.50 \times 10^{-3}\,\mathrm{m})}{(1.0 \times 10^{-3}\,\mathrm{m})^2} = \boxed{1.0\,\mathrm{mT}}. \tag{23.4.4}$$

Because B is proportional to r inside the wire, halving r halves the field.

(c) At $r = 5.0\ \mathrm{mm}$ the loop is outside the wire and encloses all $10\ \mathrm{A}$, so

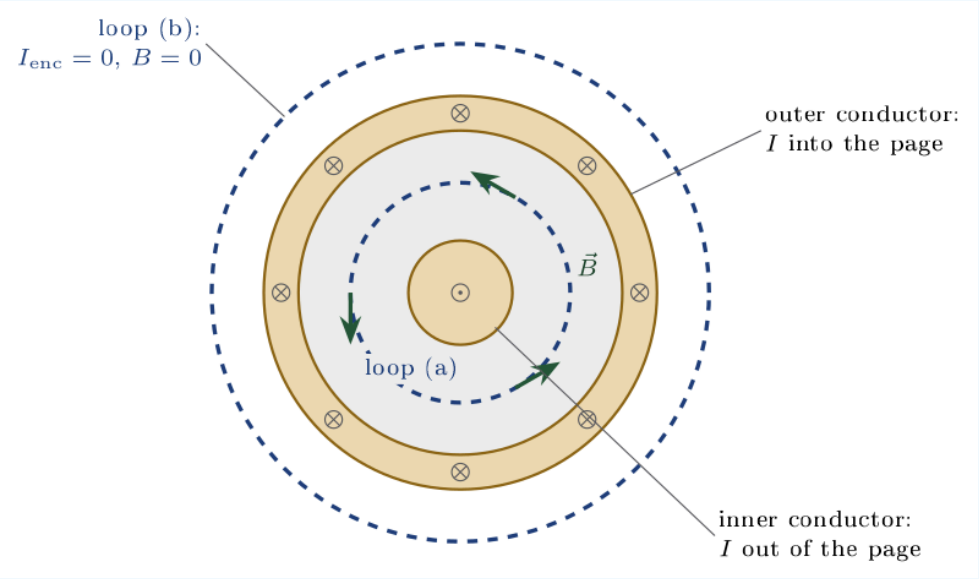
$$B = \frac{\mu_0 I}{2\pi r} = \frac{(2 \times 10^{-7}\,\mathrm{T \cdot m/A})(10\,\mathrm{A})}{5.0 \times 10^{-3}\,\mathrm{m}} = \boxed{0.40\,\mathrm{mT}}, \tag{23.4.5}$$

one-fifth of the surface value, since this point is five times as far from the axis.

Check. The field is largest right at the surface and smaller on both sides of it, just as the graph in Figure 23.8b shows.

Example 23.6: A coaxial cable

A coaxial cable has an inner conductor carrying current I one way and an outer cylindrical conductor, surrounding it, carrying the same current I back. Use Ampère's law to find the field (a) between the conductors and (b) outside the cable.



•Figure description

Cross-section of a coaxial cable. A small inner conductor at the center carries current out of the page, shown by a dot in a circle. It is surrounded by a gray insulating layer and then by a ring-shaped outer conductor carrying current into the page, shown by crosses in circles. A dashed Ampèrian loop labeled loop a runs through the insulator between the conductors, with field arrows B tangent to it, circulating counterclockwise. A second dashed loop, labeled loop b, lies outside the whole cable and is labeled enclosed current equals zero, B equals zero.

Solution

Take circular Ampèrian loops centered on the axis.

(a) Between the conductors, the loop encloses only the inner current, so $B = \mu_0 I/2\pi r$, the field of the inner wire alone.

(b) Outside the cable, the loop encloses $I - I = 0$. By symmetry B has the same magnitude all around the loop, so $B(2\pi r) = 0$ and $\boxed{B = 0}$. A coaxial cable confines its magnetic field entirely between its conductors. This is why coax is used to carry television and radio-frequency signals: its fields neither leak out to interfere with other equipment nor pick up interference from outside.

Checkpoint 23.3

Three long parallel wires pass through a closed Ampèrian loop, carrying $5\ \mathrm{A}$ upward, $3\ \mathrm{A}$ downward, and $2\ \mathrm{A}$ upward; a fourth wire, carrying $10\ \mathrm{A}$ upward, lies outside the loop. (a) What is $\oint \mathbf{B} \cdot d\mathbf{s}$ around the loop, taking upward current as positive? (b) Does the fourth wire affect the field at points on the loop? Does it affect the integral?

Answer

Answers to Checkpoints

Checkpoint 23.3. (a) Ampère's law counts only the current passing through the loop: $\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I_{\mathrm{enc}} = \mu_0(5 - 3 + 2)\ \mathrm{A} = \mu_0(4\ \mathrm{A})$. (b) The fourth wire *does* contribute to the field at every point on the loop: its field lines pass right through that region. But it does not change the integral: its field adds as much to $\oint \mathbf{B} \cdot d\mathbf{s}$ along one part of the loop as it subtracts along another, so its net contribution is zero. The field depends on all currents; the integral depends only on the enclosed ones.

This page titled 23.4: Ampère's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.5: Solenoids and Toroids

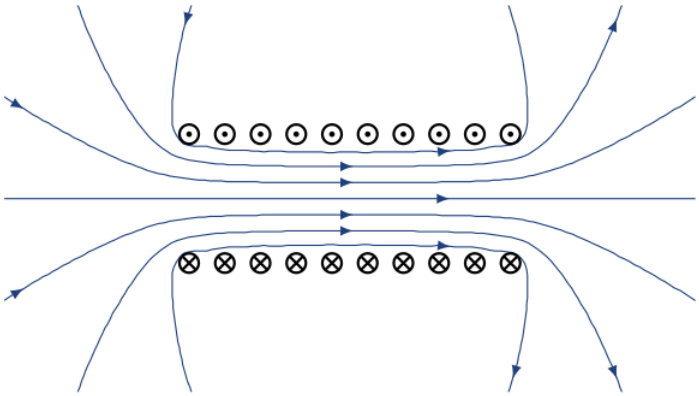


Figure 23.9. Computed field lines of a solenoid, in a plane through its axis. The current comes out of the page along the top and goes into it along the bottom. Inside, the field is strong and nearly uniform; outside it is weak; and beyond the ends it resembles the field of a bar magnet.

•Figure description

Computed magnetic field lines of a solenoid of ten turns, seen in cross-section, with its axis horizontal. The turns appear as two rows of wire cross-sections: along the top row the current comes out of the page, and along the bottom row it goes into the page. Inside the coil the field lines run straight, parallel, and evenly spaced from left to right. They emerge from the right end, spread out, and curve back around the outside of the coil, far apart, to re-enter the left end.

A **solenoid** is a long wire wound into a tight helical coil. It is the most practical way to produce a strong, uniform magnetic field, and it appears in everything from doorbells and car starters to MRI scanners. Each turn is nearly a circular loop, and the fields of all the turns add. Figure 23.9 shows the computed field lines of a solenoid of ten turns. Inside, the fields of the turns reinforce each other and the field lines are nearly straight, parallel, and evenly spaced: the field is strong and uniform. Outside, the lines spread through all the space around the coil, and the field is weak. Beyond the ends, the pattern is that of a bar magnet: a solenoid is a magnetic dipole whose north pole is the end from which the field lines emerge.

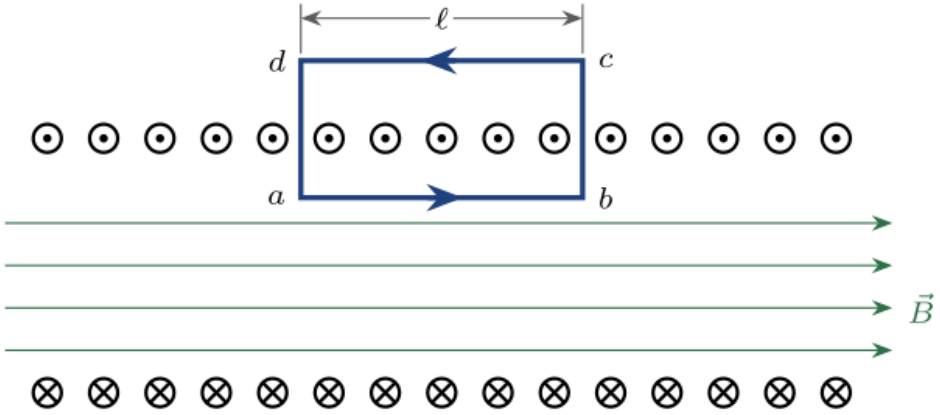


Figure 23.10. Applying Ampère's law to an ideal solenoid. Only the side ab, inside the solenoid, contributes to $\oint \mathbf{B} \cdot d\mathbf{s}$. The loop encloses $n\ell$ turns.

•Figure description

A cross-section of a long ideal solenoid with its axis horizontal. The top row of wire cross-sections carries current out of the page, and the bottom row carries current into the page. Inside the solenoid, uniform field lines B point to the right. A rectangular Ampèrian loop a b c d is drawn with its bottom side a b, of length ell, inside the solenoid along the field, its top side c d outside the solenoid above the top row of wires, and its vertical sides b c and d a crossing the top row of wires. The loop is traversed counterclockwise, from a to b to c to d.

The ideal solenoid. For a solenoid that is long compared with its diameter and tightly wound, the field inside is uniform and parallel to the axis, and the field outside is negligible. Ampère's law then gives the field inside. Take a rectangular Ampèrian loop

$abcd$ with side ab of length ℓ inside the solenoid, parallel to the axis, and side cd outside (Figure 23.10). Along ab, $\mathbf{B}$ is parallel to $d\mathbf{s}$ and contributes $B\ell$. Along bc and da, $\mathbf{B}$ is perpendicular to the path inside the solenoid and zero outside it, so they contribute nothing; and along cd, $B = 0$. So $\oint \mathbf{B} \cdot d\mathbf{s} = B\ell$. If the solenoid has n turns per unit length, the loop encloses $n\ell$ turns, each carrying current I, so $I_{\text{enc}} = n\ell I$. Ampère's law gives $B\ell = \mu_0 n\ell I$:

Key Equation: Field inside an ideal solenoid

$$B = \mu_0 n I. \tag{23.9}$$

The field does not depend on the solenoid's diameter or on the position inside it. It depends only on the current and on how closely the turns are wound. The direction inside follows the right-hand rule for loops: curl your fingers along the current in the turns, and your thumb points along $\mathbf{B}$.

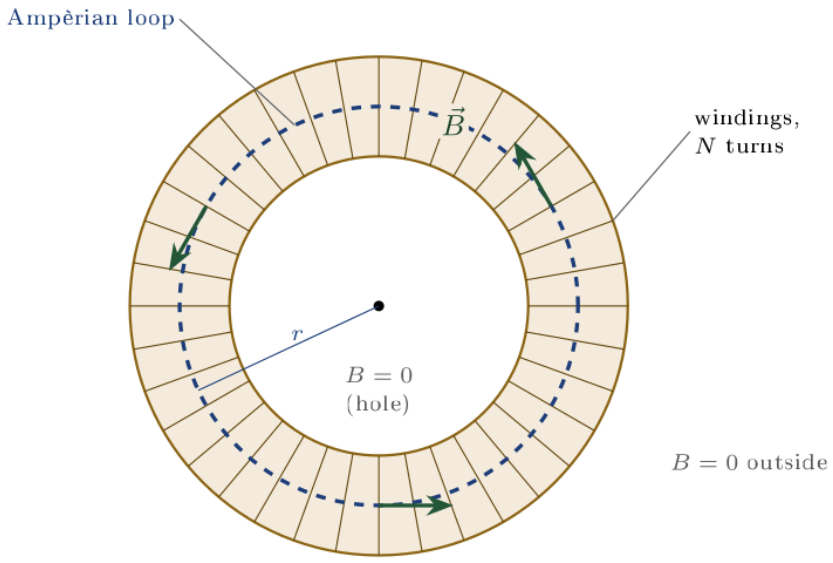


Figure 23.11. A toroid, seen from above. Inside the windings, the field lines are circles, and $B = \mu_0 NI/2\pi r$. In the central hole and outside the toroid, the field is zero.

•Figure description

A toroid seen from above: a ring-shaped core, drawn as an annulus between two circles, wound around with many turns of wire, shown as short lines crossing the annulus radially. A dashed circular Ampèrian loop of radius r runs around inside the core, with field arrows B tangent to it, circulating counterclockwise. The windings are labeled N turns. The central hole and the region outside the toroid are labeled B equals zero.

Example 23.7: An MRI magnet

The main magnet of an MRI scanner is a superconducting solenoid producing a field of $1.5\ \text{T}$. If it is wound with 2000 turns per meter, what current does it carry? What would it take to produce the same field with an ordinary copper winding?

Solution

$$I = \frac{B}{\mu_0 n} = \frac{1.5}{(4\pi \times 10^{-7})(2000)} = \boxed{600\ \text{A}}. \tag{23.5.1}$$

In a superconducting wire this current flows with no resistance at all (Chapter 20), and once established it circulates for years with the power supply disconnected. A copper winding carrying $600\ \text{A}$ through kilometers of wire would dissipate megawatts as heat, and would need a cooling plant of its own. Keeping the superconducting wire below its critical temperature is why MRI scanners are filled with liquid helium.

Toroids. Bend a solenoid around into a closed ring, like a doughnut, and it becomes a **toroid** (Figure 23.11). By symmetry, the field lines inside are circles centered on the toroid's axis. Take a circular Ampèrian loop of radius r inside the windings. It encloses all N turns, each carrying I in the same direction, so $B(2\pi r) = \mu_0 NI$, and

Key Equation: Field inside a toroid

$$B = \frac{\mu_0 NI}{2\pi r}. \tag{23.10}$$

Unlike the solenoid's, the toroid's field is not uniform: it is stronger near the inner edge. A circular loop in the central hole encloses no current, and one outside the toroid encloses each turn twice, once going up and once coming down, for a net of zero; in both regions $B = 0$. An ideal toroid confines its field completely. Toroidal coils are used in transformers and in fusion reactors called tokamaks, where the field holds a ring of hot plasma away from the walls.

Checkpoint 23.4

(a) A solenoid's length is doubled, keeping the same number of turns and the same current. What happens to the field inside? (b) What if instead the number of turns is doubled along with the length? (c) Why is the field of a toroid stronger near its inner edge than near its outer edge?

Answer

Answers to Checkpoints

Checkpoint 23.4. (a) $B = \mu_0 nI$ with $n = N/L$. Doubling L with the same N halves n, so the field inside halves. (b) Doubling both N and L leaves n unchanged, so the field is unchanged: only the number of turns per unit length matters, not the solenoid's overall size. (c) Inside a toroid, $B = \mu_0 NI/2\pi r$. Every Ampèrian circle inside encloses the same current NI, but a circle at the inner edge is shorter, so the same total $\oint \mathbf{B} \cdot d\mathbf{s}$ requires a larger B there.

This page titled 23.5: Solenoids and Toroids is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.6: Gauss's Law in Magnetism

Gauss's law for electric fields relates the electric flux through a closed surface to the charge inside it (Chapter 17). The same idea applies to the magnetic field. The *magnetic flux* through a surface is defined exactly as the electric flux was, $\Phi_B = \int \mathbf{B} \cdot d\mathbf{A}$, counting the number of field lines that cross it (Chapter 24 will make great use of it). Its SI unit is the $\mathrm{T} \cdot \mathrm{m}^2$, called the weber.

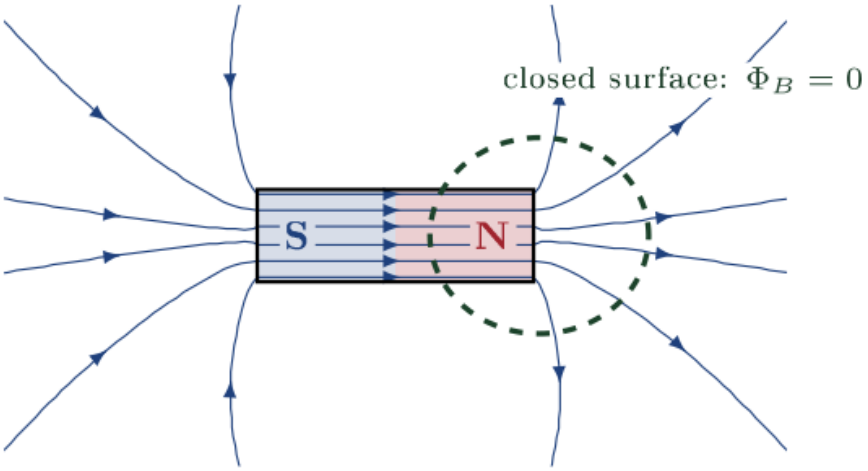


Figure 23.12. Gauss's law for magnetism. The closed surface (dashed) surrounds the north pole of a magnet, yet the net flux through it is zero. The field lines that leave through its outer part are exactly those that entered through the part inside the magnet. There is no "magnetic charge" at the pole.

•Figure description

The computed field lines of a bar magnet, with a dashed closed surface drawn as an oval enclosing the north-pole end of the magnet. Inside the magnet, field lines run from the south pole to the north pole and enter the enclosed region; they emerge from the north pole and leave the region through the curved part of the surface. Every line that enters the closed surface also leaves it, so the net magnetic flux through the surface is zero.

Magnetic field lines never begin or end: they form closed loops. So every field line that enters a closed surface must leave it again, and the total flux through any closed surface is zero (Figure 23.12). This is **Gauss's law for magnetism**:

Key Equation: Gauss's law for magnetism

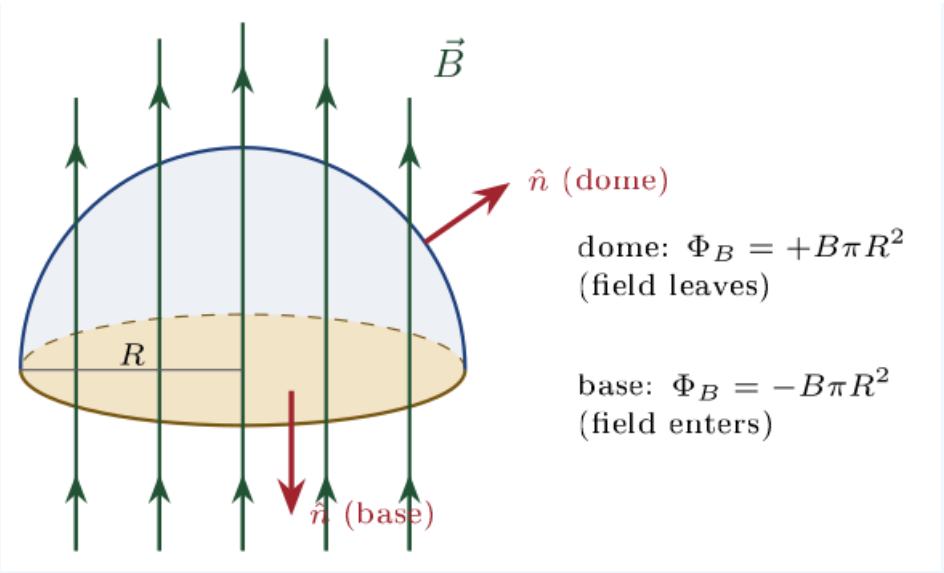
$$\oint \mathbf{B} \cdot d\mathbf{A} = 0. \tag{23.11}$$

Compare Gauss's law for electric fields, $\oint \mathbf{E} \cdot d\mathbf{A} = q_{\text{enc}}/\epsilon_0$. The magnetic version has zero on the right-hand side because there are no magnetic charges. There is no magnetic counterpart of the electric point charge, no isolated north or south pole on which field lines could begin or end. Such an object, a **magnetic monopole**, is allowed by the mathematics, and some theories of the early universe predict that monopoles should exist. In 1931 Paul Dirac showed that the existence of even one monopole anywhere would explain why electric charge is quantized. Physicists have searched for them in cosmic rays, in moon rocks, and in particle collisions, without success. As far as we know, every magnetic field is produced by moving charge, and equation (23.11) is exact.

This is why a magnet's poles cannot be separated (Section 22.1). A north pole is not a source of field lines the way a positive charge is; it is simply the place where field lines that have been running through the magnet come out of it.

Example 23.8: Flux through a dome

A uniform magnetic field $\mathbf{B}$ points vertically upward. A hemispherical dome of radius R rests on a flat circular base. What is the magnetic flux through the curved surface of the dome?



•Figure description
Side view of a hemispherical dome of radius R sitting on a flat circular base, shown in perspective as an ellipse. Five vertical field lines B point upward; each enters the closed surface through the base and leaves through the dome. The outward normal of the base points down, and the outward normal of the dome points away from its surface. Labels read: base, flux equals minus B pi R squared, field enters; dome, flux equals plus B pi R squared, field leaves.

Solution

The dome and its base together form a closed surface, so by Gauss's law the total flux through them is zero. The flux through the base, where the field enters from below, is $-B\pi R^2$ (taking the outward normal as positive). So the flux out through the dome must be $\boxed{+B\pi R^2}$, the same as the flux through a flat disk of radius R. Every field line that enters through the base leaves through the dome, and the calculation needs no integral at all.

This page titled 23.6: Gauss's Law in Magnetism is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.7: Magnetic Properties of Matter

A bar magnet has no wires and no battery. If every magnetic field comes from moving charge, where are the currents in a magnet? The answer lies in its atoms, and a full account of it requires the quantum mechanics of Part III. But the main ideas can be seen with a simple model.

Magnetism and Atoms

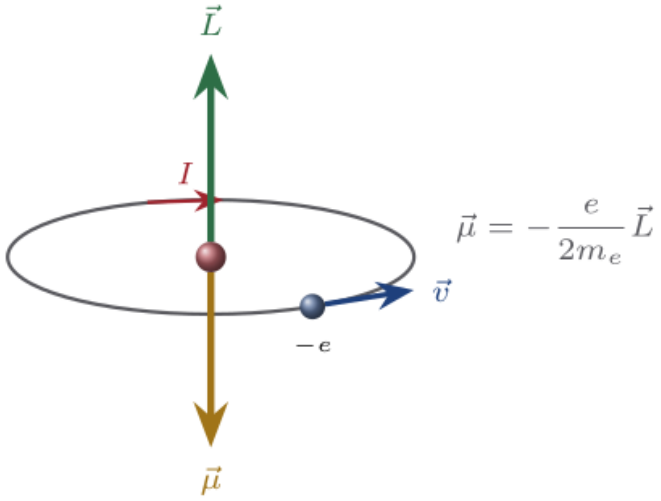


Figure 23.13. An orbiting electron is a tiny current loop. Because its charge is negative, the current circulates opposite to its motion, and its magnetic moment points opposite to its angular momentum.

•Figure description

An electron in a circular orbit around a nucleus, drawn in perspective as a horizontal ellipse. The electron moves counterclockwise when seen from above, so at the front of the orbit its velocity points to the right, while the conventional current, opposite to the electron's motion, circulates clockwise. The orbital angular momentum L points straight up from the nucleus, and the magnetic moment mu points straight down.

Orbital magnetic moment. Picture an electron moving in a circular orbit of radius r with speed v around a nucleus (Figure 23.13). It passes any point of its orbit once each period $T = 2\pi r/v$, so it constitutes a current loop with current $I = e/T = ev/2\pi r$, circulating opposite to the electron's motion because the electron is negative. The loop's magnetic moment is

$$\mu = IA = \frac{ev}{2\pi r}\,\pi r^2 = \frac{evr}{2}. \tag{23.7.1}$$

The electron's orbital angular momentum is $L = m_e v r$ (Chapter 11), so

$$\mu = -\frac{e}{2m_e}\mathbf{L}. \tag{23.12}$$

The magnetic moment is proportional to the angular momentum, and opposite to it. This relation, derived from a classical picture, survives intact in quantum mechanics. What quantum mechanics adds is that angular momentum comes in units of $\hbar = h/2\pi$ (Chapter 41). With $L = \hbar$, the moment is the natural unit of atomic magnetism, the **Bohr magneton**:

$$\mu_B = \frac{e\hbar}{2m_e} = 9.27 \times 10^{-24}\,\mathrm{A \cdot m^2}. \tag{23.13}$$

Spin magnetic moment. Electrons also have an intrinsic angular momentum, their *spin* (Chapter 41), and with it an intrinsic magnetic moment almost exactly equal to one Bohr magneton. It is not produced by any literal spinning; it is simply a property of the electron, like its charge and mass. Protons and neutrons have spin magnetic moments too, but roughly a thousand times smaller, because their mass is so much larger. Small as they are, nuclear moments are what an MRI scanner detects.

Atoms as magnets. An atom's total magnetic moment is the vector sum of the orbital and spin moments of all its electrons. In most atoms, electrons pair up with opposite spins and opposite orbital motions, and their moments cancel. Atoms with completed electron shells, such as helium and neon, have no net moment at all. Atoms with unpaired electrons, such as iron, cobalt, nickel, and the rare earths, have permanent moments of a few Bohr magnetons. In a bulk material, what matters is the **magnetization** $\mathbf{M}$, the net magnetic moment per unit volume. Its unit is $\mathrm{A \cdot m^2/m^3 = A/m}$.

Example 23.9: The magnetization of iron

Iron has $8.5 \times 10^{28}\ \mathrm{atoms/m^3}$, and each atom has a magnetic moment of about $2.2\mu_B$. If all the moments are aligned, what is the magnetization, and what field $\mu_0 M$ does it correspond to?

Solution

$$M = n\mu = (8.5 \times 10^{28})(2.2)(9.27 \times 10^{-24}) = \boxed{1.7 \times 10^6\ \mathrm{A/m}}. \tag{23.7.2}$$

The corresponding field is $\mu_0 M = (4\pi \times 10^{-7})(1.7 \times 10^6) = \boxed{2.2\ \mathrm{T}}$. This is close to the measured saturation field of iron, about $2.1\ \mathrm{T}$. A solenoid would need $nI = M = 1.7$ million ampere-turns per meter to produce the same field with no iron at all. Aligned atomic moments are an enormously concentrated source of magnetic field.

Diamagnetism vs. Paramagnetism vs. Ferromagnetism

Put a material inside a solenoid, and the field changes. If $\mathbf{B}_0$ is the field that the solenoid's current alone produces, the material contributes its own field $\mu_0\mathbf{M}$, and the total is $\mathbf{B} = \mathbf{B}_0 + \mu_0\mathbf{M}$. For most materials the magnetization is proportional to the applied field,

$$\mathbf{M} = \chi_m \frac{\mathbf{B}_0}{\mu_0}, \qquad \mathbf{B} = (1 + \chi_m)\mathbf{B}_0, \tag{23.14}$$

where the dimensionless constant χ_m is the **magnetic susceptibility**. (The combination $\kappa_m = 1 + \chi_m$ is the material's relative permeability, the magnetic counterpart of the dielectric constant of Section 19.3.) Materials fall into three classes (Table 23.1).

Table 23.1. Magnetic susceptibilities of some materials at room temperature (approximate; for ferromagnets, χ_m depends on the field and on the sample's history).

Class	Material	Susceptibility χ_m
Diamagnetic	Water	-9.0×10^{-6}
	Copper	-9.7×10^{-6}
	Bismuth	-1.7×10^{-4}
	Pyrolytic graphite	-4.5×10^{-4}
	Superconductor	-1
Paramagnetic	Oxygen gas (STP)	1.9×10^{-6}
	Aluminum	2.2×10^{-5}
	Platinum	2.7×10^{-4}
	Liquid oxygen	3.5×10^{-3}
Ferromagnetic	Nickel	about 600 (maximum)
	Iron (annealed)	about 5000 (maximum)
	Mu-metal	about 10^5 (maximum)

Diamagnetism. In a **diamagnetic** material, χ_m is small and negative: the material's magnetization opposes the applied field and weakens it slightly. Diamagnetism occurs in every material. An applied field slightly changes the orbital motion of every electron, and the change always produces a moment opposing the field (a consequence of Faraday's law, Chapter 24). In most materials

this effect is swamped by the others, but in materials whose atoms have no permanent moments, such as water, copper, and bismuth, it is all there is. A diamagnetic material is pushed *away* from a strong magnet, toward weaker field. The force is usually tiny, but in a field of $16\ \mathrm{T}$ it is enough to levitate a living frog, whose water content makes it diamagnetic, as a famous experiment demonstrated in 1997. A superconductor is a perfect diamagnet, $\chi_m = -1$: it expels magnetic fields completely, which is how a magnet can float above it.

Paramagnetism. In a **paramagnetic** material, the atoms have permanent magnetic moments, but they point in random directions. An applied field tends to line them up, since that lowers their energy $-\mu \cdot \mathbf{B}$, but thermal agitation keeps knocking them out of alignment. The result is a small net magnetization along the field: χ_m is small and positive, and a paramagnetic material is weakly attracted into a strong field. Because the competition is between alignment and thermal motion, the magnetization of a paramagnet is proportional to B_0/T: it is stronger at low temperature. This is **Curie's law**, discovered by Pierre Curie in 1895.

Example 23.10: Why paramagnetism is weak

Compare the magnetic energy of an atomic moment of one Bohr magneton in a field of $1.0\ \mathrm{T}$ with the typical thermal energy $k_B T$ at room temperature, $300\ \mathrm{K}$.

Solution

The magnetic energy is at most $\mu_B B = (9.27 \times 10^{-24})(1.0) = 9.3 \times 10^{-24}\ \mathrm{J}$. The thermal energy is $k_B T = (1.38 \times 10^{-23})(300) = 4.1 \times 10^{-21}\ \mathrm{J}$ (Chapter 36). The ratio is

$$\frac{\mu_B B}{k_B T} = \frac{9.3 \times 10^{-24}}{4.1 \times 10^{-21}} = \boxed{2.2 \times 10^{-3}}. \tag{23.7.3}$$

Thermal jostling is several hundred times more effective at scrambling the moments than the field is at aligning them, so only a fraction of a percent of the possible alignment survives. Cooling to liquid-helium temperature, $4\ \mathrm{K}$, improves the ratio some seventy-five times, as Curie's law predicts.

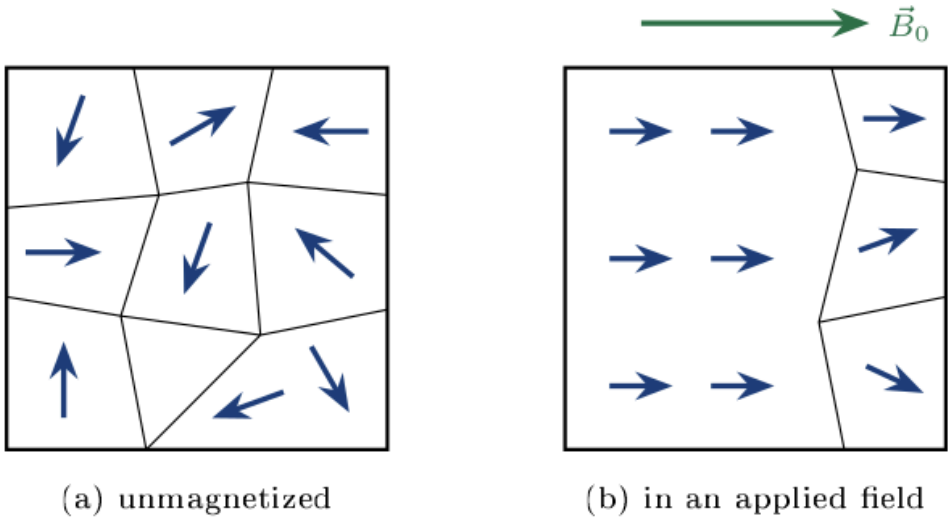


Figure 23.14. Magnetic domains. (a) In unmagnetized iron, each domain is fully magnetized, but in different directions, so the net magnetization is small. (b) In an applied field, domains aligned with the field grow, the others shrink, and the net magnetization becomes large.

•Figure description

Two panels showing magnetic domains in a piece of ferromagnetic material, drawn as a square divided into irregular regions, each containing an arrow showing its direction of magnetization. Panel a, unmagnetized: the arrows in the nine domains point in many different directions, so their magnetizations cancel. Panel b, in an applied field B zero pointing to the right: the domains magnetized to the right have grown and the others have shrunk, and most arrows now point to the right, giving a large net magnetization.

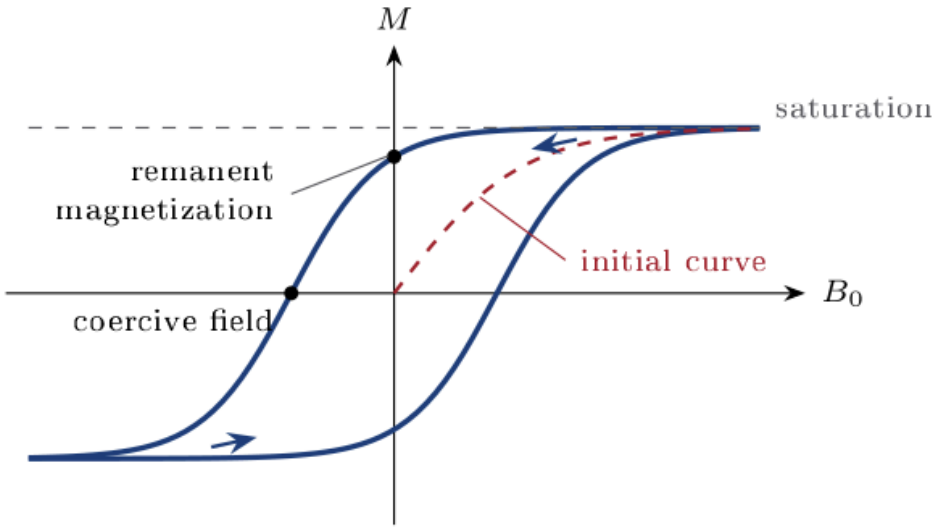


Figure 23.15. A hysteresis loop. Starting from an unmagnetized sample (dashed), the magnetization rises to saturation. When the applied field is removed, a remanent magnetization remains; a reverse, coercive field is needed to remove it. Cycling the field traces out the loop.

•Figure description

A hysteresis loop: a graph of the magnetization M of a ferromagnet against the applied field B zero. A dashed initial magnetization curve rises from the origin to saturation at the upper right. As B zero is then reduced to zero, M follows an upper branch that stays high, crossing the vertical axis at the remanent magnetization. M reaches zero only at a negative applied field, the coercive field, and saturates in the opposite direction at the lower left. Increasing B zero again brings M up along a lower branch, symmetric to the upper one, so the curve forms a closed loop.

Ferromagnetism. In iron, nickel, cobalt, and a few other materials and alloys, the magnetization is not weak at all: χ_m can be thousands. These materials are **ferromagnetic**. In them a quantum-mechanical effect, the exchange interaction, makes the moments of neighboring atoms strongly prefer to point the same way. Within regions called **magnetic domains**, typically a fraction of a millimeter across, all the moments are aligned, and each domain is magnetized to saturation. But in an unmagnetized piece of iron, the domains point in different directions and their magnetizations cancel (Figure 23.14a). An applied field makes the domains aligned with it grow at the expense of the others, as the walls between domains move, and at higher fields it rotates whole domains into line (Figure 23.14b). A modest applied field can thus produce an enormous magnetization.

Hysteresis and permanent magnets. The domain walls do not move freely; they catch on impurities and defects in the crystal. As a result, a ferromagnet's magnetization depends not only on the applied field but on its history, a behavior called **hysteresis** (Figure 23.15). When the applied field is turned back to zero, the domains do not all return, and the material keeps a *remanent* magnetization: it has become a permanent magnet. A reverse field, the *coercive* field, is needed to bring the magnetization back to zero. "Hard" magnetic materials such as neodymium–iron–boron alloys have wide hysteresis loops and make strong permanent magnets. "Soft" materials such as annealed iron have narrow loops, lose their magnetization easily, and waste little energy when magnetized back and forth, so they are used in transformer cores and electromagnets. Hysteresis is also what makes magnetic data storage possible: the bits on a hard disk are tiny regions magnetized one way or the other, which stay that way until rewritten.

The Curie temperature. Heat a ferromagnet and the thermal agitation eventually overwhelms the exchange interaction. Above a critical temperature, the **Curie temperature**, the domains disappear and the material becomes an ordinary paramagnet. For iron the Curie temperature is $1043\ \mathrm{K}$ ($770\,^\circ\mathrm{C}$), for nickel $627\ \mathrm{K}$, and for cobalt $1388\ \mathrm{K}$; for gadolinium it is just $293\ \mathrm{K}$, so a piece of gadolinium is ferromagnetic on a cool day and not on a warm one. When volcanic rock cools through its Curie temperature, its magnetic minerals freeze in the direction of Earth's field at that moment. That fossil magnetism is the record of the reversals of Earth's field mentioned in Chapter 22, and it provided some of the key evidence for plate tectonics.

Checkpoint 23.5

(a) A piece of aluminum and a piece of bismuth are held near the pole of a strong magnet. Which is attracted, and which repelled? (b) Why does hammering or heating a permanent magnet weaken it? (c) Why are transformer cores made of soft rather than hard magnetic material?

Answer

Answers to Checkpoints

Checkpoint 23.5. (a) Aluminum is paramagnetic, so it is weakly *attracted* to the pole. Bismuth is diamagnetic (strongly so, for a diamagnet), so it is weakly *repelled*. Both effects are far too small to notice without a strong magnet. (b) A permanent magnet is strong because its domains are aligned. Hammering jolts the domain walls loose so that domains can flip into random orientations, and heating adds thermal agitation that does the same; above the Curie temperature, ferromagnetism disappears altogether. (c) A transformer's field reverses direction many times a second. A soft magnetic material is easily magnetized and demagnetized, with a narrow hysteresis loop, so little energy is lost as heat in each cycle. A hard material, with its wide loop, would waste a great deal of energy.

This page titled 23.7: Magnetic Properties of Matter is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.S: Magnetic Fields and Currents (Summary)

Summary

- Magnetic fields are produced by moving charges. The **Biot–Savart law**, $d\mathbf{B} = (\mu_0/4\pi)\ I\ d\mathbf{s} \times \hat{r}/r^2$, gives the field of a current element; $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$ is the permeability of free space.
- A long straight wire produces $B = \mu_0 I/2\pi R$, in circles given by the right-hand rule. A circular arc gives $B = \mu_0 I\phi/4\pi R$ at its center, and a full loop $\mu_0 I/2R$.
- Parallel currents attract and antiparallel currents repel, with force per unit length $\mu_0 I_1 I_2/2\pi d$.
- A current loop produces the field of a **magnetic dipole**; on its axis $B = \mu_0 I R^2/2(z^2 + R^2)^{3/2}$, and far away $B = (\mu_0/4\pi)(2\mu/z^3)$, the same form as an electric dipole's field. Bar magnets and the Earth are magnetic dipoles.
- **Ampère's law**: $\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I_{\mathrm{enc}}$. With symmetry it gives the field inside a wire ($\propto r$), outside a coaxial cable (zero), inside an ideal **solenoid** ($B = \mu_0 nI$), and inside a **toroid** ($B = \mu_0 NI/2\pi r$).
- **Gauss's law for magnetism**: $\oint \mathbf{B} \cdot d\mathbf{A} = 0$. Field lines form closed loops, and there are no magnetic monopoles.
- Atoms are magnets because of the orbital and spin moments of their electrons, $\mu = -(e/2m_e)\mathbf{L}$ for orbits, in units of the **Bohr magneton** $\mu_B = e\hbar/2m_e$. **Diamagnetic** materials ($\chi_m < 0$) weakly oppose an applied field; **paramagnetic** materials ($\chi_m > 0$, small) weakly enhance it; **ferromagnetic** materials, with aligned domains, enhance it enormously, show hysteresis, and lose their ferromagnetism above the Curie temperature.

Key Equations

Biot–Savart law

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I\,d\mathbf{s} \times \hat{r}}{r^2} \tag{23.S.1}$$

Long straight wire; arc; loop center

$$B = \frac{\mu_0 I}{2\pi R}, \qquad B = \frac{\mu_0 I\phi}{4\pi R}, \qquad B = \frac{\mu_0 I}{2R} \tag{23.S.2}$$

Force between parallel currents

$$\frac{F}{L} = \frac{\mu_0 I_1 I_2}{2\pi d} \tag{23.S.3}$$

Loop axis; dipole far field

$$B = \frac{\mu_0 I R^2}{2\left(z^2 + R^2\right)^{3/2}}, \qquad B \approx \frac{\mu_0}{4\pi}\frac{2\mu}{z^3} \tag{23.S.4}$$

Ampère's law

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I_{\mathrm{enc}} \tag{23.S.5}$$

Solenoid and toroid

$$B = \mu_0 nI, \qquad B = \frac{\mu_0 NI}{2\pi r} \tag{23.S.6}$$

Gauss's law for magnetism

$$\oint \mathbf{B}\cdot d\mathbf{A} = 0 \tag{23.S.7}$$

Atomic moments; susceptibility

$$\mu = -\frac{e}{2m_e}\mathbf{L}, \qquad \mu_B = \frac{e\hbar}{2m_e}, \qquad \mathbf{B} = (1+\chi_m)\mathbf{B}_0 \tag{23.S.8}$$

This page titled 23.S: Magnetic Fields and Currents (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

23.E: Magnetic Fields and Currents (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T \cdot m/A}$.

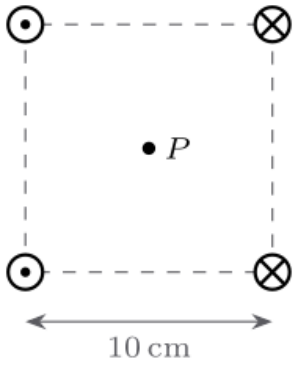
Section 23.1 — Magnetic Field Due to a Current

23.1 ★ At what distance from a long straight wire carrying $25\ \mathrm{A}$ is the magnetic field equal to Earth's field, $50\ \mu\mathrm{T}$?

23.2 ★ A circular coil of 200 turns and radius $8.0\ \mathrm{cm}$ carries $0.50\ \mathrm{A}$. Find the field at its center.

23.3 ★★ *[Real-world: power lines.]* A high-voltage line $20\ \mathrm{m}$ above the ground carries $800\ \mathrm{A}$. Find the magnetic field it produces at ground level, and compare it with Earth's field. Should people living near power lines worry that the magnetic field is strong?

23.4 ★★ Four long parallel wires pass through the corners of a square $10\ \mathrm{cm}$ on a side, as shown. Each carries $5.0\ \mathrm{A}$, in the directions indicated. Find the magnitude and direction of the field at the center of the square.



•Figure description
A square 10 centimeters on a side, with a long wire perpendicular to the page at each corner. The two wires on the left side carry current out of the page, shown as dots in circles, and the two wires on the right side carry current into the page, shown as crosses in circles. The center of the square is marked P.

23.5 ★★ Show that a straight wire segment carrying current I produces a field $B = (\mu_0 I/4\pi R)(\sin\theta_2 - \sin\theta_1)$ at a point a perpendicular distance R from its line, where θ_1 and θ_2 are the angles, measured from the perpendicular, to the two ends of the segment. Check that it gives equation (23.2) for an infinite wire.

23.6 ★★★ Use the result of Problem 23.5 to show that a square loop of side a carrying current I produces a field $B = 2\sqrt{2}\ \mu_0 I/\pi a$ at its center. Compare it with the field at the center of a circular loop of the same perimeter.

23.7 ★★★ A wire loop consists of two concentric semicircular arcs, of radii a and $b > a$, joined by two radial straight segments. It carries current I. Find the field at the common center.

Section 23.2 — Force Between Two Parallel Currents

23.8 ★ Two long parallel wires $4.0\ \mathrm{cm}$ apart carry $15\ \mathrm{A}$ and $25\ \mathrm{A}$ in the same direction. Find the force per meter on each. Do they attract or repel?

23.9 ★★ *[Conceptual.]* A loosely coiled spring hangs vertically with its bottom end dipping into a cup of mercury, completing a circuit. When a large current is switched on, the spring jumps up out of the mercury, breaking the circuit, then falls back, and the process repeats. Explain.

23.10 ★★ A long horizontal wire carries $60\ \mathrm{A}$. A second wire, parallel to it and $1.5\ \mathrm{cm}$ directly below it, has a mass of 2.0

$\mathrm{g/m}$. What current must the lower wire carry, and in which direction, for it to be suspended by the magnetic force alone? Is this equilibrium stable against vertical displacements?

23.11 ★★★ A long straight wire carries current I_1. A rectangular loop of sides a (parallel to the wire) and b, carrying current I_2, lies in the same plane with its near side a distance d from the wire. Find the net force on the loop.

Section 23.3 — Magnetic Dipoles

23.12 ★ A current loop has a magnetic moment of $0.20\ \mathrm{A \cdot m^2}$. Estimate the field it produces $50\ \mathrm{cm}$ away along its axis.

23.13 ★★ A loop of radius R carries current I. At what distance along its axis has its field fallen to one-eighth of its value at the center?

23.14 ★★ *[Real-world: geophysics.]* Use Example 23.4 to estimate the current that would have to circulate around Earth's equator, in a single loop of radius R_E, to produce Earth's dipole moment. (The real currents flow in the core, of radius about $3500\ \mathrm{km}$.)

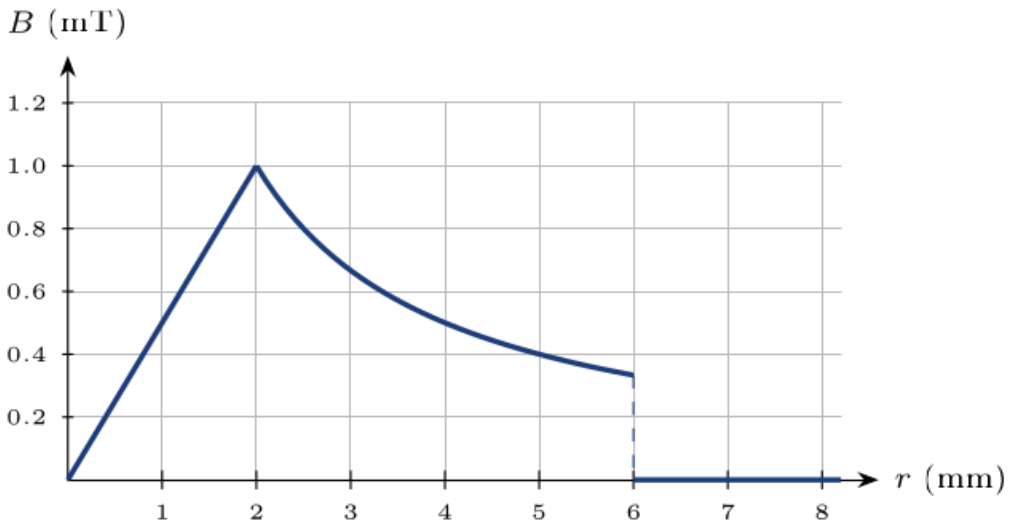
23.15 ★★★ *[Real-world: laboratory fields.]* A Helmholtz coil pair consists of two identical coaxial loops of radius R, each carrying current I in the same direction, separated by a distance R. Using equation (23.6), find the field at the midpoint, and show that its second derivative along the axis is zero there, so that the field is very nearly uniform near the center.

Section 23.4 — Ampère's Law

23.16 ★ The line integral $\oint \mathbf{B} \cdot d\mathbf{s}$ around a closed path is $3.8 \times 10^{-5}\ \mathrm{T \cdot m}$. What net current passes through the path?

23.17 ★★ *[Conceptual.]* Ampère's law holds for a loop around a single short segment of current, but it cannot be used to find that segment's field. Why not? When is Ampère's law useful for finding $\mathbf{B}$?

23.18 ★★ *[Graph-reading.]* The graph shows the measured magnetic field as a function of distance r from the axis of a coaxial cable. The inner conductor is a solid wire, and the outer conductor is a thin cylindrical shell.



•Figure description
A graph of magnetic field B in millitesla, from 0 to 1.2, against distance r from the axis in millimeters, from 0 to 8. B rises linearly from zero at the axis to 1.0 millitesla at r equals 2 millimeters, then falls off as one over r, reaching about 0.33 millitesla at 6 millimeters. At 6 millimeters it drops abruptly to zero and stays zero beyond.

a. What are the radii of the inner conductor and of the outer shell?
b. Find the current in the inner conductor.
c. What is the current in the outer shell? How do you know?

23.19 ★★ A long hollow cylindrical conductor of inner radius a and outer radius b carries current I uniformly distributed over its cross-section. Find B for $r < a$, $a < r < b$, and $r > b$, and sketch $B(r)$.

23.20 ★★★ A long straight wire of radius R carries a current whose density increases with distance from the axis, $J = J_0 r/R$. Find the total current, and the magnetic field inside and outside the wire.

Section 23.5 — Solenoids and Toroids

23.21 ★ A solenoid $30\ \text{cm}$ long has 600 turns and carries $2.5\ \text{A}$. Find the field inside.

23.22 ★★ You have $20\ \text{m}$ of wire and want to make a solenoid $25\ \text{cm}$ long with a diameter of $2.0\ \text{cm}$. If the wire can safely carry $3.0\ \text{A}$, what is the strongest field you can produce inside?

23.23 ★★ A toroid has 400 turns, an inner radius of $8.0\ \text{cm}$, and an outer radius of $12\ \text{cm}$. With a current of $1.5\ \text{A}$, find the field at the inner and outer edges of the windings. By what percentage is the field at the outer edge smaller than at the inner edge?

23.24 ★★★ *[Real-world: fusion.]* A tokamak's toroidal field coils produce $5.3\ \text{T}$ at a distance of $6.2\ \text{m}$ from the torus axis (the ITER design). Find the total current NI in all the coils together. What field does this give at $4.0\ \text{m}$ and at $8.4\ \text{m}$ from the axis?

Section 23.6 — Gauss's Law in Magnetism

23.25 ★ The magnetic flux into a closed surface through one of its faces is $3.0\ \text{mWb}$, and $1.2\ \text{mWb}$ leaves through another face. What is the net flux through all the remaining faces?

23.26 ★★ *[Conceptual.]* A student claims that a closed surface drawn tightly around the north pole of a bar magnet, but not around the south pole, has a net outward magnetic flux, "because field lines come out of the north pole." Explain the error, using Figure 23.12.

23.27 ★★ A uniform field of $0.30\ \text{T}$ points along $+x$. Find the magnetic flux through each face of a cube of side $10\ \text{cm}$ with its edges along the axes, and verify that the total is zero.

Section 23.7.1 — Magnetism and Atoms

23.28 ★ In the Bohr model, the electron in hydrogen's ground state moves at $2.19 \times 10^{6}\ \text{m/s}$ in a circle of radius $5.29 \times 10^{-11}\ \text{m}$. Find the current and the magnetic moment of the orbit, and compare with the Bohr magneton.

23.29 ★★ A cobalt atom has a magnetic moment of about $1.7\mu_B$, and cobalt has $9.0 \times 10^{28}\ \text{atoms/m}^3$. Find its saturation magnetization and the corresponding field $\mu_0 M$.

Section 23.7.2 — Diamagnetism vs. Paramagnetism vs. Ferromagnetism

23.30 ★ A solenoid produces $0.20\ \text{mT}$ when empty. Find the field when it is filled with (a) aluminum and (b) annealed iron with $\chi_m = 5000$. (c) Why would the formula fail for iron if the empty solenoid produced $0.010\ \text{T}$?

23.31 ★★ *[Conceptual.]* Liquid oxygen poured between the poles of a strong magnet is caught and held there, while liquid nitrogen runs straight through. What does this tell you about the O_2 and N_2 molecules?

23.32 ★★ The magnetization of a paramagnetic salt is $120\ \text{A/m}$ at $300\ \text{K}$ in a certain field. Assuming Curie's law, what is it in the same field at $4.0\ \text{K}$?

23.33 ★★★ Show that the maximum energy of a magnetic moment μ in a field B is comparable to $k_B T$ when $B \approx k_B T/\mu$. Find this field for a moment of one Bohr magneton at $300\ \text{K}$ and at $1.0\ \text{K}$, and explain why low-temperature paramagnets can be almost fully magnetized in laboratory fields.

This page titled 23.E: Magnetic Fields and Currents (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24: FARADAY'S LAW

Oersted had shown that a current produces a magnetic field. Surely, thought Michael Faraday, a magnetic field could produce a current. For ten years he tried, winding coils around magnets and placing wires beside steady currents, and saw nothing. Then, in August 1831, he wound two coils on opposite sides of an iron ring, connected one to a battery and the other to a galvanometer, and noticed that the galvanometer needle jumped at the instant he closed the battery circuit, and jumped the other way when he opened it. While the current flowed steadily, nothing happened. The effect came not from the magnetic field, but from its *change*.

That discovery, electromagnetic induction, powers the modern world. Nearly all of the world's electrical energy is produced by generators that work on Faraday's principle. Transformers, electric guitars, induction cooktops, wireless chargers, and the card readers that power a contactless card's chip when you tap to pay all depend on it. It also completes the picture of the fields themselves: a changing magnetic field creates an electric field. This chapter defines magnetic flux, states Faraday's law and Lenz's law for the induced emf and its direction, and applies them to conductors moving through fields and to generators and motors.

Learning Objectives

- Calculate the magnetic flux through a surface.
- Use Faraday's law to find the emf induced by a changing magnetic flux, and describe the induced electric field.
- Use Lenz's law to find the direction of an induced current, and relate it to energy conservation.
- Find the motional emf of a conductor moving through a magnetic field, and account for the energy involved.
- Explain how generators produce alternating emfs, and how back emf governs the current in a motor.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

24.1: Magnetic Flux
24.2: Faraday's Law of Induction
24.3: Motional EMF
24.4: Generators and Motors
24.S: Faraday's Law (Summary)
24.E: Faraday's Law (Problems)

This page titled 24: Faraday's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.1: Magnetic Flux

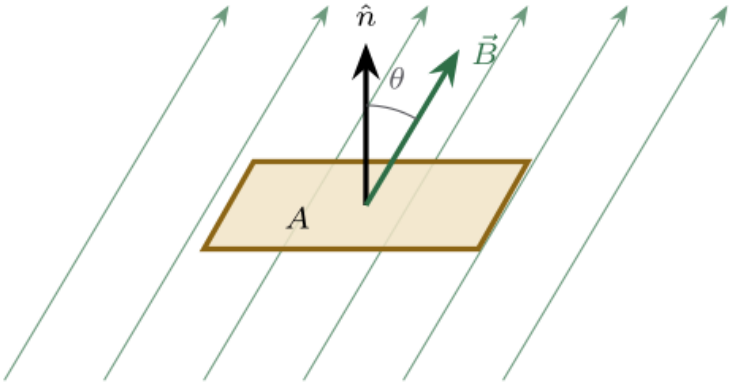


Figure 24.1. The magnetic flux through a flat loop in a uniform field is $\Phi_B = BA\cos\theta$, where θ is the angle between $\mathbf{B}$ and the normal $\hat{n}$ to the loop.

•Figure description

A flat rectangular loop of area A, drawn in perspective and tilted, in a uniform magnetic field B represented by parallel arrows pointing up and to the right. A unit vector n-hat, normal to the loop, makes an angle theta with the direction of B. Only the component of B along n-hat, B cos theta, passes through the loop.

Faraday found that the induced effects depend on how many magnetic field lines pass through a circuit, and on how fast that number changes. The quantity that measures this is the magnetic flux, defined exactly as the electric flux was in Chapter 17. For a flat surface of area A in a uniform magnetic field, let $\hat{n}$ be a unit vector normal to the surface, and θ the angle between $\mathbf{B}$ and $\hat{n}$ (Figure 24.1). The **magnetic flux** through the surface is

$$\Phi_B = \mathbf{B}\cdot\mathbf{A} = BA\cos\theta, \tag{24.1}$$

where $\mathbf{A} = A\hat{n}$. It is largest when the field passes straight through the surface ($\theta = 0$), and zero when the field lies parallel to the surface ($\theta = 90^\circ$), so that no field lines cross it. If the field is not uniform or the surface is not flat, divide the surface into small elements $d\mathbf{A}$ and add their contributions:

Key Equation: Magnetic flux

$$\Phi_B = \int \mathbf{B}\cdot d\mathbf{A}. \tag{24.2}$$

The SI unit of magnetic flux is the **weber** (Wb), after the German physicist Wilhelm Weber: $1\ \text{Wb} = 1\ \text{T}\cdot\text{m}^2$. For a surface that is not closed, the direction of $\hat{n}$ is a choice, and it sets the sign of the flux. For a surface bounded by a circuit, we will tie it to the direction around the circuit by the right-hand rule. For a *closed* surface, the flux is always zero, by Gauss's law for magnetism (Section 23.6).

Example 24.1: Flux through a loop in Earth's field

A horizontal circular loop of radius $20\ \text{cm}$ lies on a table where Earth's field is $50\ \mu\text{T}$, pointing down into the ground at 60° below the horizontal. Find the flux through the loop.

Solution

Take $\hat{n}$ pointing straight down. The field makes an angle of $90^\circ - 60^\circ = 30^\circ$ with the vertical, so

$$\Phi_B = BA\cos\theta = (5.0\times10^{-5}\ \text{T})\,\pi(0.20\,\text{m})^2\cos 30^\circ = \boxed{5.4\times10^{-6}\ \text{Wb}}. \tag{24.1.1}$$

Only the vertical component of Earth's field contributes: the horizontal component lies in the plane of the loop.

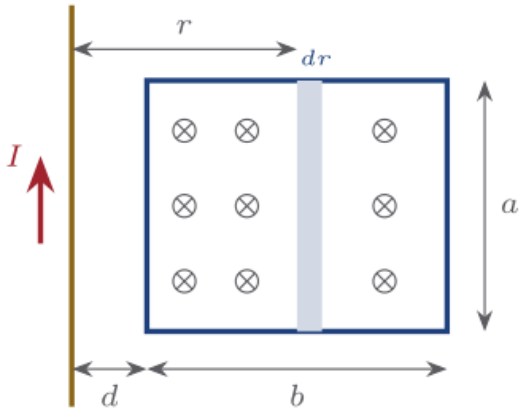


Figure 24.2. Finding the flux of a long wire's field through a rectangular loop. The field is uniform along each thin strip parallel to the wire, so the flux is found by adding the strips.

•Figure description

A long vertical straight wire carrying current I upward, and to its right, in the same plane, a rectangular loop with its long sides of length a parallel to the wire. The near side of the loop is a distance d from the wire and the loop has width b. A thin vertical strip of the loop, of width d r, is shaded at a distance r from the wire. The wire's magnetic field points into the page over the loop, shown by crosses in circles.

Example 24.2: Flux beside a long wire

A long straight wire carries $20\ \text{A}$. A rectangular loop of length $a = 10\ \text{cm}$ (parallel to the wire) and width $b = 4.0\ \text{cm}$ lies in the same plane, with its near side $d = 1.0\ \text{cm}$ from the wire (Figure 24.2). Find the flux through the loop.

Solution

The field of the wire, $\mu_0 I/2\pi r$, is perpendicular to the plane of the loop but not uniform: it weakens with distance r from the wire. Divide the loop into thin strips of width dr and length a, parallel to the wire. The field is uniform across each strip, so its flux is $d\Phi_B = B\ a\ dr$, and

$$\begin{aligned}\Phi_B &= \int_d^{d+b} \frac{\mu_0 I}{2\pi r}\, a\, dr = \frac{\mu_0 I a}{2\pi} \ln\frac{d+b}{d} \\ &= (2\times 10^{-7})(20)(0.10)\ln 5.0 = \boxed{6.4\times 10^{-7}\ \text{Wb}}.\end{aligned} \tag{24.1.2}$$

This page titled 24.1: Magnetic Flux is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.2: Faraday's Law of Induction

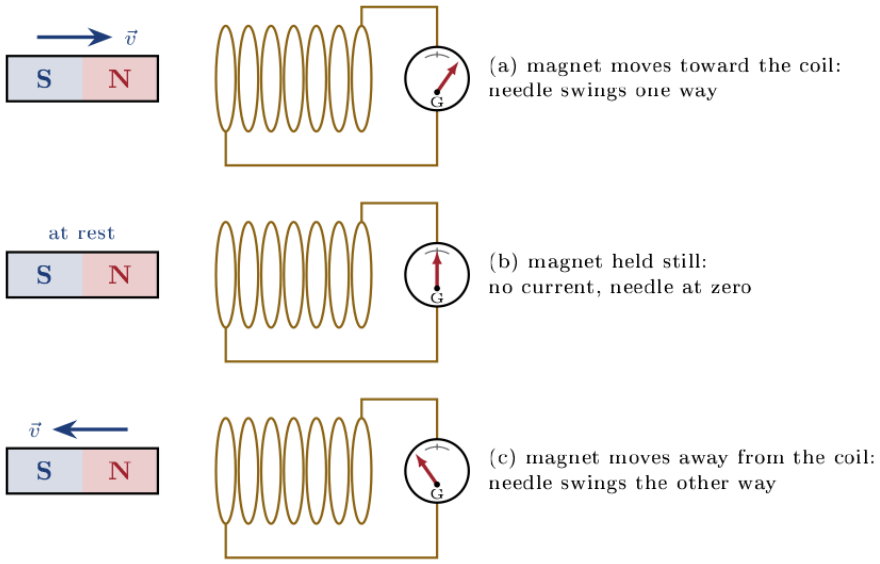


Figure 24.3. Faraday's experiment. (a) Moving a magnet toward a coil induces a current, shown by the galvanometer (G). (b) While the magnet is held still, the flux through the coil does not change and no current flows. (c) Moving the magnet away reverses the current.

•Figure description

Faraday's experiment in three panels, one above another. In each, a bar magnet with its north pole facing right sits to the left of a coil of wire whose axis is horizontal, and the two ends of the coil are wired to a galvanometer, a meter drawn as a circle with a needle and a zero mark at the top. Panel a: the magnet moves to the right, toward the coil, with velocity v, and the needle swings to the right. Panel b: the magnet is at rest, and the needle points straight up at zero. Panel c: the magnet moves to the left, away from the coil, and the needle swings to the left.

Faraday's experiments, and many like them, can all be summarized in one statement. Push a bar magnet into a coil connected to a galvanometer, and a current flows while the magnet moves (Figure 24.3). Hold the magnet still, and the current stops. Pull it out, and the current flows the other way. Move the coil instead of the magnet, and the result is the same. Replace the magnet with a second coil, and switch its current on or off, and again a current flows in the first coil, but only while the current in the second is changing. In every case, a current is induced whenever the magnetic flux through the circuit changes, and the faster it changes, the larger the current. This is **electromagnetic induction**.

What is induced, more fundamentally, is an emf, which drives the current through the circuit's resistance. **Faraday's law of induction** states that the **induced emf** in a circuit equals the negative of the rate of change of the magnetic flux through it:

Key Equation: Faraday's law

$$\mathcal{E} = -\frac{d\Phi_B}{dt} \qquad (\text{for a coil of } N \text{ turns, } \mathcal{E} = -N\frac{d\Phi_B}{dt}). \tag{24.3}$$

For a tightly wound coil, the same flux passes through each of the N turns, and their emfs add, like batteries in series. The minus sign comes from Lenz's law: the induced emf always acts to oppose the change in flux that produces it. Section 24.2 explains what that means for the direction of the emf. The emf is in volts, since $1\ \text{Wb/s} = 1\ \text{T}\cdot\text{m}^2/\text{s} = 1\ \text{V}$. If the circuit has resistance R, the induced current is $I = |\mathcal{E}|/R$.

Since $\Phi_B = BA\cos\theta$ for a flat loop in a uniform field, there are three ways to induce an emf: change the field B; change the area A of the loop inside the field; or change the angle θ by turning the loop. The first is how transformers work (Chapter 25), the second is the motional emf of Section 24.3, and the third is how generators work (Section 24.4).

Example 24.3: A coil in a changing field

A coil of 200 turns, each of area $4.0 \times 10^{-3}\ \mathrm{m}^2$, has its plane perpendicular to a uniform magnetic field. The field increases steadily from $0.10\ \mathrm{T}$ to $0.50\ \mathrm{T}$ in $0.20\ \mathrm{s}$. (a) Find the magnitude of the induced emf. (b) If the coil has a resistance of $8.0\ \Omega$, what current flows, and how much charge passes through it?

Solution

(a) The flux through each turn changes by $\Delta\Phi_B = A\ \Delta B = (4.0 \times 10^{-3})(0.40) = 1.6 \times 10^{-3}\ \mathrm{Wb}$, at a steady rate. So

$$|\mathcal{E}| = N\frac{\Delta\Phi_B}{\Delta t} = (200)\,\frac{1.6 \times 10^{-3}\,\mathrm{Wb}}{0.20\,\mathrm{s}} = \boxed{1.6\,\mathrm{V}}. \tag{24.2.1}$$

(b) $I = |\mathcal{E}|/R = 1.6/8.0 = \boxed{0.20\,\mathrm{A}}$, flowing for $0.20\ \mathrm{s}$, so $Q = I\ \Delta t = \boxed{0.040\,\mathrm{C}}$. Notice that $Q = N\ \Delta\Phi_B/R$ does not depend on how fast the field changed: a faster change gives a larger current for a shorter time. Faraday measured flux this way, from the charge that passed through a galvanometer.

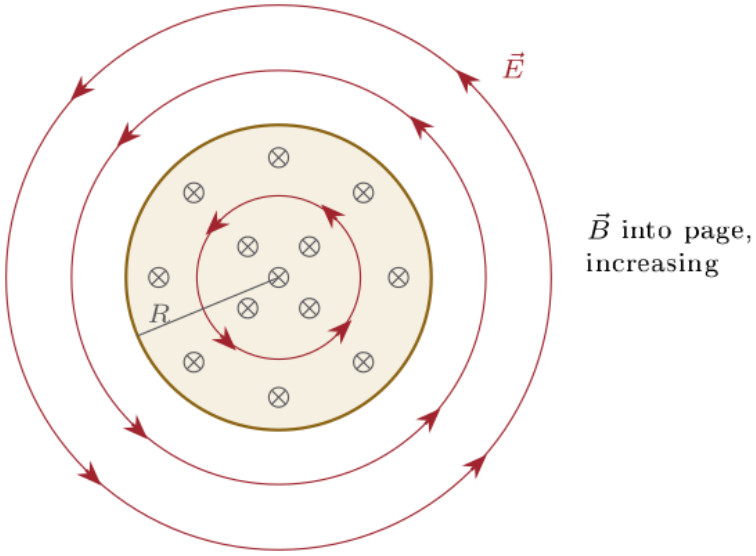


Figure 24.4. A magnetic field that is changing inside a solenoid induces an electric field whose field lines are closed circles, both inside and outside. For a field into the page that is increasing, the induced field circulates counterclockwise.

•Figure description

The cross-section of a long solenoid, a circle of radius R, seen end-on, with its magnetic field into the page, shown by crosses in circles, increasing with time. Circular induced electric field lines, centered on the axis, circulate counterclockwise both inside and outside the solenoid. A dashed circle of radius r inside the solenoid is marked, with E tangent to it.

The induced electric field. What pushes the charges around the circuit? The circuit need not move, so it is not a magnetic force. It must be an electric field, and Faraday's law says where it comes from: a changing magnetic flux produces an electric field. The emf around a closed loop is the work per unit charge done by this field, $\oint \mathbf{E}\cdot d\mathbf{s}$, so Faraday's law can be written in terms of the fields alone:

$$\oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}. \tag{24.4}$$

The loop need not be a wire. This **induced electric field** exists in empty space wherever the magnetic field changes, whether or not there is a conductor to carry a current. It is quite unlike the electric field of static charges. Its field lines do not begin and end on charges; they form closed loops around the changing magnetic field (Figure 24.4). And its line integral around a closed path is not zero, so it is *not* conservative: it cannot be described by an electric potential, and Kirchhoff's loop rule does not hold for a path around a changing flux (Section 21.2). Equation (24.4) is one of Maxwell's four equations (Chapter 27).

Example 24.4: Stimulating the brain

In transcranial magnetic stimulation (TMS), a coil held against the head produces a brief pulse of magnetic field. Inside the head, the field is roughly uniform over a region of radius about $2\ \text{cm}$ and rises at a rate of about $1.0 \times 10^4\ \text{T/s}$. Estimate the induced electric field $1.0\ \text{cm}$ from the axis of the field region.

Solution

By symmetry, the induced field lines are circles about the axis, and E is the same all around a circle of radius r. Apply equation (24.4) to such a circle, inside the field region, where the flux it encloses is $\pi r^2 B$:

$$E(2\pi r) = \pi r^2 \frac{dB}{dt} \qquad \Longrightarrow \qquad E = \frac{r}{2}\frac{dB}{dt} = \frac{0.010\,\text{m}}{2}\,(1.0 \times 10^4\,\text{T/s}) = \boxed{50\,\text{V/m}}. \tag{24.2.2}$$

(We have dropped the minus sign, which only sets the direction.) A field of this size in the brain tissue drives currents strong enough to trigger nerve impulses, without any electrodes touching the patient. TMS is used to map brain function and to treat depression.

Note: Induction at work

An electric guitar's pickup is a small coil wound around a magnet. The magnet magnetizes the steel string above it, and as the string vibrates, the flux through the coil oscillates at the string's frequency, inducing an emf that is amplified to make the sound. A ground-fault circuit interrupter, the outlet with test and reset buttons in bathrooms and kitchens, surrounds both wires of the circuit with a small coil. Normally the currents in the two wires are equal and opposite and their fields cancel; if some current leaks to ground, perhaps through a person, the imbalance produces a changing flux, the induced emf trips a switch, and the power is cut in milliseconds. Induction cooktops, wireless phone chargers, and the readers for transit cards all work by induction too.

Lenz's Law

The minus sign in Faraday's law gives the direction of the induced emf, but it is usually easier to find the direction with a rule stated by Heinrich Lenz in 1834. **Lenz's law**: *the induced current flows in the direction that makes its own magnetic field oppose the change in flux that produced it.*

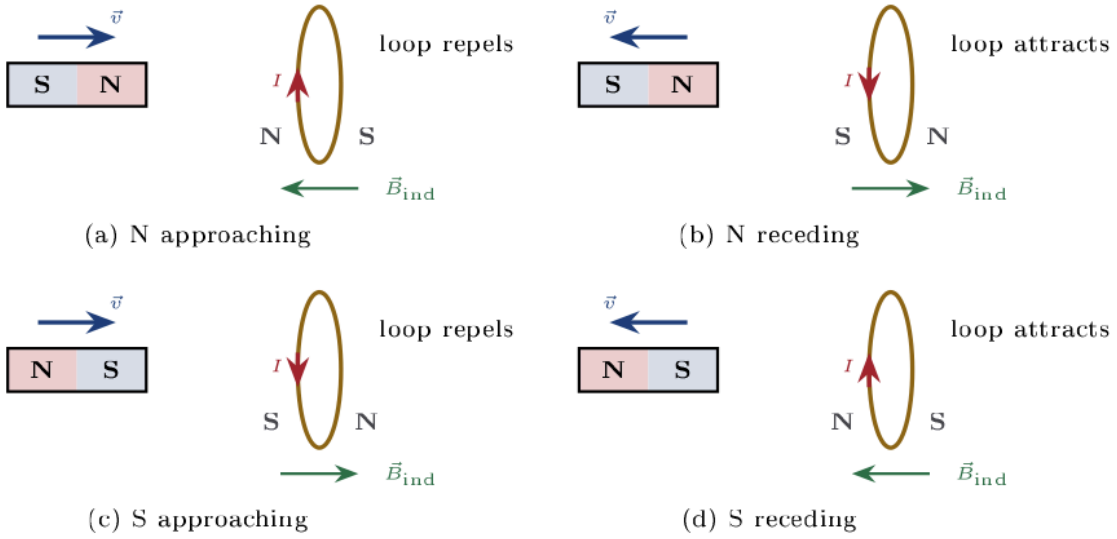


Figure 24.5. Lenz's law. In each case the induced current in the loop makes a field $\mathbf{B}_{\text{ind}}$ that opposes the change in flux, so that the loop repels an approaching magnet and attracts a receding one. The arrow shows the current on the front side of the loop. The letters beside each loop give the poles of the loop's own field: the loop acts like a small bar magnet, with the face nearest the magnet the same pole as the approaching end (so it repels) or the opposite pole of the receding end (so it attracts).

•Figure description

Four panels, each showing a bar magnet on the left and a circular wire loop on the right, seen in perspective with its axis horizontal. Panel a: the north pole faces the loop and moves toward it; the flux through the loop to the right increases, and the induced current creates a field pointing left, so the loop behaves like a small magnet whose face nearest the bar magnet is a north pole, marked N beside the loop, and it repels the magnet. Panel b: the north pole moves away; the flux decreases and the induced current creates a field to the right, so the loop's face nearest the magnet becomes a south pole, marked S, and attracts the magnet. Panel c: the south pole faces the loop and moves toward it; the induced field points right and the loop's near face becomes a south pole, repelling. Panel d: the south pole moves away; the induced field points left and the loop's near face becomes a north pole, attracting. In each panel an arrow on the front of the loop shows the direction of the induced current.

Note that the induced current opposes the *change* in the flux, not the flux itself. Figure 24.5 shows four cases. When the north pole of a magnet approaches a loop (a), the flux through the loop, directed away from the magnet, increases. The induced current makes a field pointing back toward the magnet, to oppose the increase; the face of the loop toward the magnet acts as a north pole, and repels the magnet. When the north pole recedes (b), the flux decreases, and the induced current makes a field in the *same* direction as the magnet's, to prop up the flux; now the near face of the loop acts as a south pole, and attracts the receding magnet. Reversing the magnet (c, d) reverses everything. Every time, the induced current pushes against the motion that caused it.

Lenz's law and energy. Lenz's law is a consequence of energy conservation. Suppose the induced current flowed the other way, so that an approaching north pole met a south pole and was pulled in. The magnet would speed up, increasing the induced current, which would pull it in harder: it would gain kinetic energy while the current delivered energy to the circuit's resistance, all for nothing. Because the induced current always opposes the motion, you must do work to move the magnet, and that work is exactly the electrical energy that appears in the circuit.

Eddy currents. Induced currents flow not only in wires but in any piece of metal where the flux is changing. These swirling **eddy currents** dissipate energy as heat, and by Lenz's law they oppose the motion that produces them. A metal plate swinging into a strong magnetic field stops abruptly, as if it had entered thick syrup. This **magnetic braking** is used to stop roller coasters and high-speed trains smoothly, with no contact and no wear. Eddy currents are also what heat the pan on an induction cooktop. Where they are unwanted, as in the iron cores of transformers and motors, the cores are made of thin insulated sheets, which cut the eddy currents' paths.

Example 24.5: A magnet dropped through a copper pipe

A strong magnet dropped down a vertical copper pipe falls far more slowly than a stone dropped beside it, even though copper is not attracted to magnets. Explain.

Solution

As the magnet falls, the flux through each ring of the pipe below it increases, and the flux through each ring above it decreases. By Lenz's law, the rings below carry eddy currents that repel the approaching magnet, and the rings above carry currents that attract the departing magnet, pulling it back. Both forces point up. The faster the magnet falls, the faster the flux changes, and the larger the currents and the upward force. The magnet quickly reaches a terminal speed at which the magnetic drag balances its weight, much like an object falling through a fluid (Section 6.2). The magnet's lost gravitational potential energy becomes heat in the pipe.

Checkpoint 24.1

A circular loop lies flat on a table. A magnet held above it, with its north pole down, is lifted straight up and away. (a) Does the flux through the loop increase or decrease? (b) Seen from above, does the induced current circulate clockwise or counterclockwise? (c) Is the magnet attracted or repelled by the loop as it rises?

Answer

Answers to Checkpoints

Checkpoint 24.1. (a) Near the north pole the field points downward through the loop, so the downward flux *decreases* as the magnet moves away. (b) By Lenz's law the induced current opposes the change, so it tries to keep the downward flux from falling: it produces its own downward field inside the loop. Seen from above, a current making a downward field circulates *clockwise*. (c) Attracted. The loop now acts like a magnet whose south pole faces up, toward the magnet's north pole, so it pulls back on the rising magnet: opposing the motion that causes the change, as Lenz's law requires.

This page titled 24.2: Faraday's Law of Induction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.3: Motional EMF

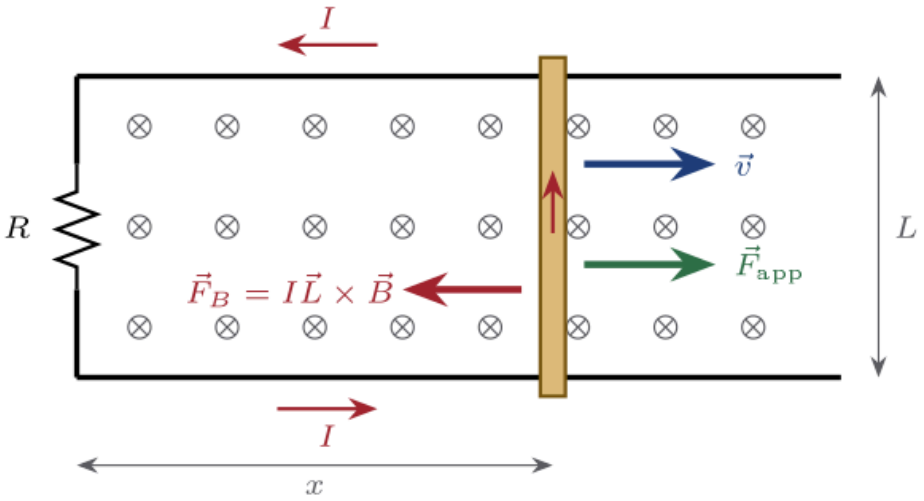


Figure 24.6. A rod sliding on rails in a magnetic field into the page. The motional emf BLv drives a counterclockwise current. The magnetic force on the current in the rod opposes its motion, so an applied force is needed to keep it moving.

•Figure description

A conducting rod of length L lies across two long horizontal parallel rails. The rails are joined on the left by a resistor R. The whole arrangement lies in a uniform magnetic field pointing into the page, shown by crosses in circles. The rod moves to the right with velocity v, pulled by an applied force. The induced current I flows up through the rod, left along the top rail, down through the resistor, and right along the bottom rail: counterclockwise. The magnetic force on the current in the rod, I L cross B, points to the left, opposing the motion.

An emf is also induced when a conductor moves through a magnetic field. Consider a conducting rod of length L sliding with speed v along two parallel conducting rails that are joined at one end by a resistor R, in a uniform field $\mathbf{B}$ perpendicular to the plane of the rails (Figure 24.6).

From the magnetic force. The free charges in the rod move with it, with velocity $\mathbf{v}$, so each feels a magnetic force $q\mathbf{v} \times \mathbf{B}$, directed along the rod. For positive carriers, with $\mathbf{B}$ into the page and $\mathbf{v}$ to the right, the force points up the rod. The charges are pushed along the rod, doing work qvB per unit charge per unit length, or $qvBL$ along its whole length. The rod acts as a battery with emf

Key Equation: Motional emf

$$\mathcal{E} = BLv. \tag{24.5}$$

This is **motional emf**. It requires $\mathbf{B}$, $\mathbf{v}$, and the rod all to be perpendicular to each other; if they are not, only the perpendicular components count.

From Faraday's law. The same result follows from Faraday's law. The rod, the rails, and the resistor form a closed circuit. As the rod moves, the area inside the circuit grows: if the rod is a distance x from the resistor, the flux is $\Phi_B = BLx$, and it changes at the rate $d\Phi_B/dt = BL\ dx/dt = BLv$. So $|\mathcal{E}| = BLv$, as before. Motional emf is Faraday's law applied to a circuit whose area changes.

Direction and energy. As the rod moves to the right, the flux into the page increases. By Lenz's law, the induced current makes a field out of the page inside the circuit, so it circulates counterclockwise: up through the rod, as the magnetic force on positive carriers also says. The current $I = BLv/R$ in the rod feels a magnetic force $I\mathbf{L} \times \mathbf{B}$ (equation (22.8)), of magnitude

$$F_B = ILB = \frac{B^2L^2v}{R}, \tag{24.3.1}$$

pointing to the *left*, against the motion. To keep the rod moving at constant speed, you must push it with an equal force, doing work at the rate

$$P = F_{\text{app}} v = \frac{B^2 L^2 v^2}{R}. \tag{24.3.2}$$

This is exactly the rate at which the resistor dissipates energy, $I^2 R = (BLv/R)^2 R = B^2 L^2 v^2 / R$. Your mechanical work becomes electrical energy, and then heat, just as Lenz's law requires. The sliding rod is the simplest possible electric generator.

Example 24.6: A sliding rod

In Figure 24.6, $B = 0.50\ \text{T}$, $L = 0.20\ \text{m}$, and $R = 0.40\ \Omega$; the rod and rails have negligible resistance. The rod moves at $4.0\ \text{m/s}$. Find (a) the emf, (b) the current, (c) the force needed to keep the rod moving, and (d) the power delivered to the resistor.

Solution

(a) $\mathcal{E} = BLv = (0.50)(0.20)(4.0) = \boxed{0.40\,\text{V}}$.

(b) $I = \mathcal{E}/R = 0.40/0.40 = \boxed{1.0\,\text{A}}$, counterclockwise.

(c) $F = ILB = (1.0)(0.20)(0.50) = \boxed{0.10\,\text{N}}$, to the right, balancing the magnetic force.

(d) $P = I^2 R = (1.0)^2(0.40) = \boxed{0.40\,\text{W}}$. As a check, the applied force does work at the rate $Fv = (0.10)(4.0) = 0.40\ \text{W}$, the same.

Example 24.7: A flying airplane

An airliner with a wingspan of $60\ \text{m}$ flies horizontally at $250\ \text{m/s}$ where the vertical component of Earth's field is $40\ \mu\text{T}$. Find the emf between its wingtips. Could it be used to power anything?

Solution

The wing is a conductor moving perpendicular to the vertical component of the field, so

$$\mathcal{E} = BLv = (4.0 \times 10^{-5})(60)(250) = \boxed{0.60\,\text{V}}. \tag{24.3.3}$$

It cannot drive a current through a circuit on board. Any wire connected between the wingtips moves along with the wing, through the same field, and has the same emf induced in it, so the two emfs cancel around the loop. Only a circuit with part of it at rest in the field, like the rails in Figure 24.6, can deliver motional emf as a current. (Tethered satellites, trailing a conducting cable kilometers long through the ionosphere, have tested this idea in orbit, using the surrounding plasma to close the circuit.)

Checkpoint 24.2

In the sliding-rod circuit of Figure 24.6, the rod is pushed along the rails at twice its former speed. By what factor does each of these change: (a) the induced emf, (b) the current, (c) the magnetic force on the rod, and (d) the power you must supply to keep it moving at that speed?

Answer

Answers to Checkpoints

Checkpoint 24.2. (a) The emf $\mathcal{E} = BLv$ is proportional to the speed, so it doubles. (b) The circuit's resistance is unchanged, so $I = \mathcal{E}/R$ doubles too. (c) The force on the rod is ILB, proportional to the current, so it doubles. (d) To keep the rod moving at constant speed you must push with a force equal to that magnetic force, and the power is force times speed. Both have doubled, so the power rises by a factor of 4, and all of it appears as heat, $I^2 R$, in the resistance.

This page titled 24.3: Motional EMF is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.4: Generators and Motors

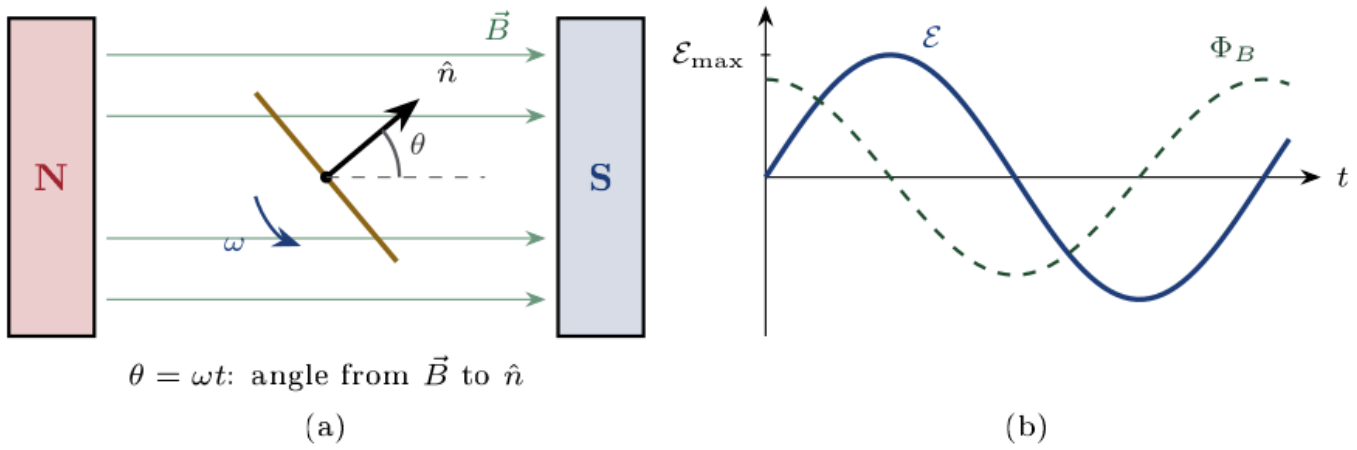


Figure 24.7. An ac generator. (a) A coil rotates in a uniform field, so the angle between its normal and the field is $\theta = \omega t$. (b) The flux (dashed) varies as $\cos \omega t$, and the induced emf (solid) as $\sin \omega t$, with amplitude $NBA\omega$.

•Figure description

Panel a: a generator coil seen end-on, looking along its axis of rotation, between the north pole of a magnet on the left and the south pole on the right, with a uniform field B pointing from left to right. The coil appears as a line segment through the axis, rotating counterclockwise with angular velocity omega; its normal n-hat makes angle theta equal to omega t with B. Panel b: a graph against time of the flux through the coil, a cosine curve, drawn dashed, and the induced emf, a sine curve of amplitude script E max, drawn solid. The emf is largest where the flux crosses zero, and zero where the flux is largest.

The ac generator. An electric **generator** converts mechanical energy into electrical energy. In its simplest form, a coil of N turns and area A rotates with constant angular velocity ω in a uniform magnetic field $\mathbf{B}$, about an axis perpendicular to the field (Figure 24.7). If the normal to the coil makes angle $\theta = \omega t$ with the field, the flux through each turn is

$$\Phi_B = BA \cos \omega t, \tag{24.4.1}$$

and Faraday's law gives an emf

$$\mathcal{E} = -N \frac{d\Phi_B}{dt} = NBA\omega \sin \omega t. \tag{24.4.2}$$

The emf oscillates sinusoidally:

Key Equation: Emf of an ac generator

$$\mathcal{E} = \mathcal{E}_{\max} \sin \omega t, \qquad \mathcal{E}_{\max} = NBA\omega. \tag{24.6}$$

The emf is largest when the plane of the coil is parallel to the field and the flux through it is zero, because that is when the flux is changing fastest; it is zero when the coil faces the field and the flux is largest. The ends of the coil are connected to the external circuit through *slip rings*, metal rings on the shaft that turn with the coil, and carbon *brushes* that press against them. The current in the circuit reverses direction every half turn. This is *alternating current* (ac), the subject of Chapter 26. In North America the generators in power stations turn so that $f = \omega/2\pi = 60$ Hz; in most of the rest of the world, 50 Hz. To produce direct current instead, the slip rings are replaced by a split-ring commutator, like the one in a dc motor (Section 22.3), which reverses the connections every half turn so that the output never changes sign.

Example 24.8: A power-station generator in miniature

A generator coil has 100 turns, each of area $0.050\ \text{m}^2$, and rotates at $60\ \text{rev/s}$ in a field of $0.30\ \text{T}$. Find the maximum emf. At what angle between the coil's normal and the field is the emf half its maximum?

Solution

With $\omega = 2\pi(60) = 377\ \text{rad/s}$,

$$\mathcal{E}_{\max} = NBA\omega = (100)(0.30)(0.050)(377) = \boxed{570\ \text{V}}. \tag{24.4.3}$$

The emf is half its maximum when $\sin\omega t = \frac{1}{2}$, at $\theta = \boxed{30^\circ}$ (and at 150°, 210°, and 330°). Commercial generators produce much larger emfs, tens of kilovolts, by using many coils, iron cores, and strong electromagnets. The work to turn them comes from steam turbines driven by burning fuel or nuclear fission, from falling water, or from the wind.

Work in a generator. When the generator delivers current to a load, the current in the coil feels a torque in the field (Section 22.3), and by Lenz's law that torque opposes the rotation. The more current is drawn, the harder the generator is to turn. That is why a bicycle dynamo makes pedaling harder only when the lamp is on, and why power stations must burn more fuel when demand rises. Electric and hybrid cars use the same effect in reverse: in *regenerative braking*, the drive motor is switched to act as a generator, and the opposing torque slows the car while the induced current recharges the battery.

Motors and back emf. An electric motor (Section 22.3) is a generator run backward: a current in the coil produces a torque that turns it. But as the coil turns in the field, it is also a generator, and an emf is induced in it. By Lenz's law, this induced emf opposes the current that is driving the motor. It is called the **back emf**, $\mathcal{E}_{\text{back}}$, and it is proportional to the motor's speed. If the motor is connected to a source of voltage V, and its coil has resistance R, the loop rule gives the current:

$$I = \frac{V - \mathcal{E}_{\text{back}}}{R}. \tag{24.7}$$

When the motor is first switched on and is not yet turning, there is no back emf, and the current is large, V/R. As the motor speeds up, the back emf grows and the current falls. A motor running freely with no load spins up until the back emf nearly equals V, drawing little current. When a load is applied, the motor slows, the back emf drops, and more current flows, supplying the extra power. The power converted to mechanical work is $\mathcal{E}_{\text{back}}I$, and I^2R is lost as heat in the windings.

Example 24.9: Back emf in a motor

A $120\ \text{V}$ dc motor has windings with a resistance of $2.0\ \Omega$. Running at full speed, it draws $4.0\ \text{A}$. Find (a) the back emf at full speed, (b) the mechanical power it delivers, and (c) the current at the instant it is switched on.

Solution

(a) From equation (24.7), $\mathcal{E}_{\text{back}} = V - IR = 120 - (4.0)(2.0) = \boxed{112\ \text{V}}$.

(b) The power delivered by the source is $VI = 480\ \text{W}$. Of this, $I^2R = 32\ \text{W}$ heats the windings, and the rest, $\mathcal{E}_{\text{back}}I = (112)(4.0) = \boxed{448\ \text{W}}$, is converted to mechanical work.

(c) At start-up the back emf is zero, so $I = V/R = 120/2.0 = \boxed{60\ \text{A}}$, fifteen times the running current. This surge is why lights sometimes dim briefly when a large motor, such as a refrigerator compressor or a power saw, starts. If the motor is overloaded and jams, the current stays near this value, and the windings can overheat and burn out.

Checkpoint 24.3

(a) The rotation rate of a generator is doubled. What happens to its maximum emf and to the frequency of its output? (b) Why does the current drawn by a blender's motor increase when a thick mixture slows its blades, and why can a motor that stalls completely overheat?

Answer

Answers to Checkpoints

Checkpoint 24.3. (a) The maximum emf, $NBA\omega$, is proportional to the angular speed, so it doubles. The frequency, the number of rotations per second, doubles as well. (b) A spinning motor acts as a generator, producing a back emf proportional to its speed that opposes the applied voltage; the current is $I = (V - \mathcal{E}_{\text{back}})/R$. When a thick mixture slows the blades, the back emf drops, the net voltage across the windings rises, and the current increases. A stalled motor has no back emf at all, so it draws the full V/R, typically many times its running current. That current heats the windings rapidly, which is why a motor that stalls can burn out.

This page titled 24.4: Generators and Motors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.S: Faraday's Law (Summary)

Summary

- The **magnetic flux** through a surface is $\Phi_B = \int \mathbf{B} \cdot d\mathbf{A}$, or $BA\cos\theta$ for a flat surface in a uniform field. Its unit is the weber, $1\ \text{Wb} = 1\ \text{T} \cdot \text{m}^2$.
- **Faraday's law**: a changing magnetic flux induces an emf $\mathcal{E} = -N\ d\Phi_B/dt$ in a circuit. The flux can change because B, the area, or the orientation changes.
- A changing magnetic field produces an **induced electric field**, with closed field lines, satisfying $\oint \mathbf{E} \cdot d\mathbf{s} = -d\Phi_B/dt$. It is not conservative.
- **Lenz's law**: the induced current opposes the change in flux that produces it. This is required by energy conservation, and it is the origin of eddy-current braking.
- A conductor of length L moving with speed v perpendicular to a field B has a **motional emf** $\mathcal{E} = BLv$. The work done to keep it moving equals the electrical energy delivered.
- A coil rotating in a field is an ac **generator**, with $\mathcal{E} = NBA\omega\sin\omega t$. A motor produces a **back emf** that limits its current, $I = (V - \mathcal{E}_{\text{back}})/R$, so a motor draws its largest current when starting.

Key Equations

Magnetic flux

$$\Phi_B = \int \mathbf{B}\cdot d\mathbf{A}, \qquad \Phi_B = BA\cos\theta \ \ (\text{uniform } B, \text{ flat surface}) \tag{24.S.1}$$

Faraday's law

$$\mathcal{E} = -N\frac{d\Phi_B}{dt}, \qquad \oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt} \tag{24.S.2}$$

Motional emf

$$\mathcal{E} = BLv \tag{24.S.3}$$

AC generator

$$\mathcal{E} = NBA\omega\sin\omega t \tag{24.S.4}$$

Motor with back emf

$$I = \frac{V - \mathcal{E}_{\text{back}}}{R} \tag{24.S.5}$$

This page titled 24.S: Faraday's Law (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

24.E: Faraday's Law (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T\cdot m/A}$.

Section 24.1 — Magnetic Flux

24.1 ★ A square loop $15\ \mathrm{cm}$ on a side is in a uniform field of $0.40\ \mathrm{T}$. Find the flux through it if the field makes an angle of (a) 0°, (b) 60°, (c) 90° with the loop's normal.

24.2 ★ A solenoid of radius $2.0\ \mathrm{cm}$ has 1000 turns per meter and carries $3.0\ \mathrm{A}$. Find the flux through one of its turns.

24.3 ★★ *[Conceptual.]* A loop lies in a uniform field with the field in the plane of the loop, so the flux is zero. Can you make the flux nonzero by (a) moving the loop sideways, (b) stretching it, (c) rotating it? Explain.

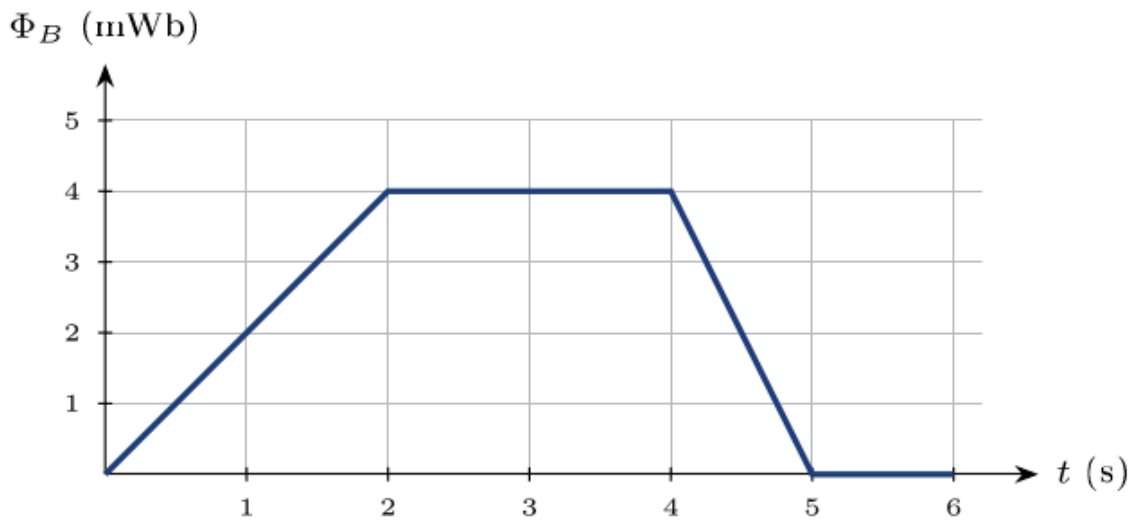
24.4 ★★★ A long straight wire carries current I along the axis of a toroid of rectangular cross-section, with inner radius a, outer radius b, and height h. Find the flux of the wire's field through the cross-section of the toroid.

Section 24.2 — Faraday's Law of Induction

24.5 ★ The flux through a coil of 50 turns changes steadily from $2.0\ \mathrm{mWb}$ to $8.0\ \mathrm{mWb}$ in $0.050\ \mathrm{s}$. Find the magnitude of the induced emf.

24.6 ★ A circular loop of radius $5.0\ \mathrm{cm}$ is perpendicular to a field that decreases from $0.80\ \mathrm{T}$ to zero in $0.10\ \mathrm{s}$. Find the average induced emf.

24.7 ★★ *[Graph-reading.]* The graph shows the magnetic flux through each turn of a 50-turn coil as a function of time.



•Figure description
A graph of magnetic flux in milliwebers, from 0 to 5, against time in seconds, from 0 to 6. The flux rises linearly from 0 at t equals 0 to 4 milliwebers at t equals 2 seconds, stays constant at 4 milliwebers until t equals 4 seconds, then falls linearly to zero at t equals 5 seconds, and stays at zero until 6 seconds.

a. Find the induced emf in each time interval, with its sign.
b. Sketch the emf as a function of time.
c. When is the emf largest in magnitude? Why?

24.8 ★★ A 20-turn coil of area $0.010\ \mathrm{m}^2$ and resistance $0.50\ \Omega$ lies perpendicular to a field $B(t) = (0.20\ \mathrm{T/s})\ t + (0.050\ \mathrm{T/s}^2)\ t^2$. Find the induced current at $t = 2.0\ \mathrm{s}$.

24.9 ★★ *[Real-world: measurement.]* A "search coil" of 100 turns and area $2.0\ \mathrm{cm}^2$, with total circuit resistance $50\ \Omega$, is placed with its plane perpendicular to an unknown field and then quickly pulled out of it. A charge of $8.0\ \mu\mathrm{C}$ passes through the circuit. What was the field? (See Example 24.3.)

24.10 ★★ The field inside a long solenoid of radius $3.0\ \mathrm{cm}$ increases at $0.50\ \mathrm{T/s}$. Find the induced electric field at distances of $1.5\ \mathrm{cm}$, $3.0\ \mathrm{cm}$, and $6.0\ \mathrm{cm}$ from the axis.

24.11 ★★★ Show that outside a long solenoid of radius R whose field changes at the rate dB/dt, the induced electric field at distance r from the axis is $E = (R^2/2r)\ dB/dt$, even though $\mathbf{B}$ itself is zero there. Sketch $E(r)$ for all r.

Section 24.2.1 — Lenz's Law

24.12 ★ *[Conceptual.]* A bar magnet is dropped, north pole down, through a horizontal wire loop. Seen from above, what is the direction of the induced current (a) as the magnet approaches the loop, and (b) after it has passed through and is falling away below it?

24.13 ★★ *[Conceptual.]* A rectangular loop lies in the plane of the page beside a long straight wire carrying a current to the right, with the loop below the wire. Find the direction of the induced current in the loop if (a) the current in the wire increases, (b) the loop moves away from the wire, (c) the loop moves parallel to the wire.

24.14 ★★ *[Conceptual.]* An aluminum ring rests on top of a vertical coil wrapped around an iron core. When the coil is connected to a battery, the ring leaps into the air. Explain, using Lenz's law. Would the ring jump if it had a narrow cut through it?

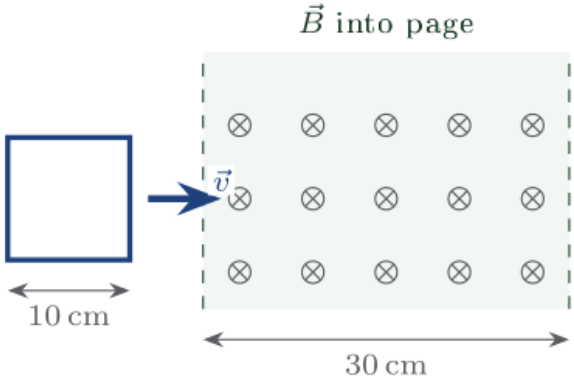
24.15 ★★ *[Real-world: brakes.]* Explain why eddy-current brakes are ineffective for bringing a train to a complete stop, and why trains that use them also carry friction brakes.

Section 24.3 — Motional EMF

24.16 ★ A car's $1.5\ \mathrm{m}$ steel axle moves at $30\ \mathrm{m/s}$ where the vertical component of Earth's field is $45\ \mu\mathrm{T}$. Find the emf between the ends of the axle.

24.17 ★★ In the sliding-rod arrangement of Figure 24.6, $B = 0.80\ \mathrm{T}$, $L = 0.50\ \mathrm{m}$, and $R = 2.0\ \Omega$. What force is needed to move the rod at a steady $3.0\ \mathrm{m/s}$, and what power is dissipated?

24.18 ★★ A square loop of side $10\ \mathrm{cm}$ and resistance $0.20\ \Omega$ moves at a constant $2.0\ \mathrm{m/s}$ through a region $30\ \mathrm{cm}$ wide containing a uniform field of $0.50\ \mathrm{T}$ into the page, as shown. Plot the induced emf and the current, with their directions, as functions of the position of the loop's leading edge, from before it enters the field to after it has left.



•Figure description
A square wire loop 10 centimeters on a side, on the left, moves to the right with velocity v toward a region 30 centimeters wide containing a uniform magnetic field into the page, shown by crosses in circles. Outside the region there is no field. The position x of the loop's leading edge is measured from the left boundary of the field region.

24.19 ★★★ A metal rod of length L rotates with angular velocity ω about one end, in a plane perpendicular to a uniform field B. Show that the emf between its ends is $\frac{1}{2}B\omega L^2$. *(Hint: each small piece dr of the rod moves with speed ωr.)*

24.20 ★★★ In the sliding-rod arrangement, the rod (mass m) is given an initial speed v_0 and then released, with no applied force. Show that its speed decreases exponentially, $v = v_0 e^{-t/\tau}$ with $\tau = mR/B^2L^2$, and that the total heat produced in the resistor equals the rod's initial kinetic energy.

Section 24.4 — Generators and Motors

24.21 ★ A generator coil of 250 turns and area $0.020\ \text{m}^2$ turns at $50\ \text{Hz}$ in a $0.15\ \text{T}$ field. Find its maximum emf.

24.22 ★★ You want to build a generator whose maximum emf is $170\ \text{V}$ at $60\ \text{Hz}$, using a field of $0.25\ \text{T}$ and a coil of area $0.012\ \text{m}^2$. How many turns does it need?

24.23 ★★ A dc motor connected to $24\ \text{V}$ has a winding resistance of $1.5\ \Omega$. At its operating speed it draws $2.0\ \text{A}$. Find (a) the back emf, (b) the mechanical power output, (c) the efficiency, and (d) the starting current.

24.24 ★★ *[Real-world: energy.]* A wind turbine's generator produces an average of $2.0\ \text{MW}$. Explain, using Lenz's law, why the turbine's blades turn more slowly when the grid draws more power from it, and where the energy comes from.

24.25 ★★★ Show that the average power delivered by an ac generator of maximum emf $\mathcal{E}_{\max}$ to a resistor R is $\mathcal{E}_{\max}^2/2R$, and that the average torque needed to turn it at angular speed ω is $\mathcal{E}_{\max}^2/2R\omega$. (Use the fact that the average of $\sin^2\omega t$ over a cycle is $\frac{1}{2}$.)

This page titled 24.E: Faraday's Law (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25: INDUCTANCE

Try to stop a current suddenly, by pulling a plug from its socket while a vacuum cleaner is running, and you may see a spark jump between the prongs. A current that has built up in a coil does not want to stop, much as a moving car does not want to stop. The coil fights the change with an induced emf, large enough, if the current is interrupted quickly, to drive a spark across an air gap. This electrical “inertia” is inductance.

Faraday’s law says that a changing current in one circuit induces an emf in any circuit whose flux it changes, and that includes the circuit itself. This chapter follows that idea. It introduces inductors, the circuit elements designed to have inductance, and the mutual inductance that couples one circuit to another. It shows that an inductor stores energy in its magnetic field, just as a capacitor stores energy in its electric field. It follows the rise and fall of current in circuits with inductors and resistors, a close cousin of the RC circuits of Chapter 21. And it ends with the transformer, the device that makes it practical to send electrical power across a continent.

Learning Objectives

- Define self-inductance and mutual inductance, and calculate them for solenoids.
- Find the emf across an inductor from the rate of change of its current.
- Calculate the energy stored in an inductor and the energy density of a magnetic field.
- Analyze the growth and decay of current in RL circuits, using the time constant L/R.
- Explain how transformers change voltages and currents, and why power is transmitted at high voltage.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

25.1: Inductors
25.2: Energy in a Magnetic Field
25.3: RL Circuits
25.4: Transformers
25.S: Inductance (Summary)
25.E: Inductance (Problems)

This page titled 25: Inductance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.1: Inductors

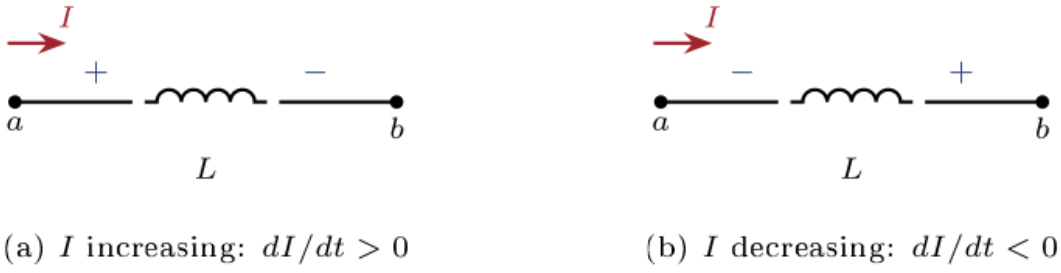


Figure 25.1. The self-induced emf of an inductor always opposes the change in current. (a) When the current is increasing, the end where it enters is at the higher potential, so the inductor opposes the increase. (b) When it is decreasing, the polarity reverses, and the inductor pushes to keep the current flowing.

•Figure description

Two panels, each showing an inductor, drawn as a coil symbol, between a point a on the left and a point b on the right, carrying current I from a to b. In panel a the current is increasing: the self-induced emf opposes the increase, so the left end a, where the current enters, is at the higher potential, marked plus, and b is marked minus. In panel b the current is decreasing: the induced emf tries to keep the current flowing, so the right end b is at the higher potential, marked plus, and a is marked minus.

A current in a coil produces a magnetic field, and that field passes through the coil's own turns. If the current changes, so does this flux, and by Faraday's law an emf is induced in the coil itself. This **self-induced emf** opposes the change in current (Lenz's law): if the current is increasing, the emf pushes against it; if it is decreasing, the emf pushes to keep it flowing (Figure 25.1).

For a coil of N turns with flux Φ_B through each turn, the flux is proportional to the current that produces it, so we can write $N\Phi_B = LI$. The constant of proportionality,

Key Equation: Self-inductance

$$L = \frac{N\Phi_B}{I}, \tag{25.1}$$

is the coil's **self-inductance**, or simply its inductance. Like capacitance, it depends only on the geometry of the coil (and on any magnetic material inside it), not on the current. Faraday's law then gives the self-induced emf:

Key Equation: Emf of an inductor

$$\mathcal{E}_L = -N\frac{d\Phi_B}{dt} = -L\frac{dI}{dt}. \tag{25.2}$$

The SI unit of inductance is the **henry** (H), after the American physicist Joseph Henry, who discovered self-induction at about the same time as Faraday discovered induction: $1\ \mathrm{H} = 1\ \mathrm{Wb/A} = 1\ \mathrm{V \cdot s/A}$. A coil with an inductance of one henry produces an emf of one volt when its current changes at one ampere per second. Practical inductors range from nanohenries in radio circuits to several henries in the power supplies of large equipment.

A circuit element designed for its inductance is an **inductor**, drawn as a coil in circuit diagrams (Figure 25.1). Every circuit has some inductance, since every current makes a magnetic field; but in most circuits it is too small to matter unless the current changes very rapidly.

Inductors in circuits. To use the loop rule (Section 21.2) with an inductor, we need the potential change across it. Crossing an inductor in the direction of the current, the potential changes by

$$\Delta V = -L\frac{dI}{dt}. \tag{25.3}$$

This is a drop when the current is increasing and a rise when it is decreasing, as Figure 25.1 shows. For a steady current, $dI/dt = 0$, and an ideal inductor (one with no resistance in its wire) has no potential difference across it at all. An inductor resists *changes* in current, not current itself. It plays the same role for current that mass plays for velocity: it does not prevent motion, but it resists acceleration.

The inductance of a solenoid. For a long solenoid of N turns, length ℓ, and cross-sectional area A, the field inside is $B = \mu_0 nI = \mu_0 NI/\ell$ (equation (23.9)), and the flux through each turn is $\Phi_B = BA = \mu_0 NIA/\ell$. So

$$L = \frac{N\Phi_B}{I} = \frac{\mu_0 N^2 A}{\ell} = \mu_0 n^2 A\ell. \tag{25.4}$$

The inductance grows as the *square* of the number of turns: doubling the turns doubles both the field and the number of turns that the flux links. Filling the solenoid with iron multiplies L by the relative permeability $1 + \chi_m$ (Section 23.7), which can be thousands, so practical inductors of large inductance are wound on iron or ferrite cores.

Example 25.1: The inductance of a solenoid

A solenoid $25\ \text{cm}$ long and $2.0\ \text{cm}$ in diameter is wound with 500 turns, with no core. (a) Find its inductance. (b) What emf is induced in it when its current rises steadily from zero to $2.0\ \text{A}$ in $1.0\ \text{ms}$?

Solution

(a) The cross-sectional area is $A = \pi(0.010)^2 = 3.14 \times 10^{-4}\ \text{m}^2$, so

$$L = \frac{\mu_0 N^2 A}{\ell} = \frac{(4\pi \times 10^{-7})(500)^2(3.14 \times 10^{-4})}{0.25} = \boxed{0.39\,\text{mH}}. \tag{25.1.1}$$

(b) The current changes at $dI/dt = (2.0\ \text{A})/(1.0 \times 10^{-3}\ \text{s}) = 2000\ \text{A/s}$, so

$$|\mathcal{E}_L| = L\frac{dI}{dt} = (3.9 \times 10^{-4}\,\text{H})(2000\,\text{A/s}) = \boxed{0.79\,\text{V}}. \tag{25.1.2}$$

While the current rises, this emf opposes it. Once the current is steady, the emf is zero.

Mutual vs. Self-Inductance

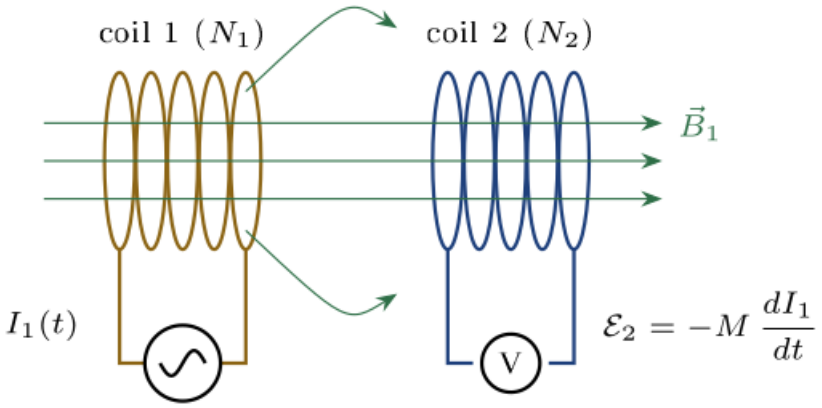


Figure 25.2. Mutual inductance. Part of the field of coil 1 passes through coil 2, so a changing current in coil 1 induces an emf in coil 2, although no wire connects them.

•Figure description

Two coaxial coils side by side. Coil 1, on the left, carries a changing current I sub 1 supplied by a source. Its magnetic field lines pass through its own turns and continue through the turns of coil 2, on the right. Coil 2 is connected to a meter and has an emf script E sub 2 induced in it. No wire connects the two coils.

Faraday's first discovery of induction used two coils: a changing current in one induced an emf in the other. When two circuits are close enough that the field of each passes through the other, a changing current in either one induces an emf in both. The flux

through coil 2 due to the current I_1 in coil 1 is proportional to I_1, and we write $N_2\Phi_{21} = MI_1$ (Figure 25.2). Faraday's law then gives the emf induced in coil 2:

Key Equation: Mutual inductance

$$M = \frac{N_2\Phi_{21}}{I_1}, \qquad \mathcal{E}_2 = -M\frac{dI_1}{dt}. \tag{25.5}$$

The constant M is the **mutual inductance** of the two coils, also measured in henries. It depends on their sizes, their numbers of turns, and how they are placed, and it is largest when the coils share a common iron core. A remarkable result, which can be proved from the Biot–Savart law, is that the mutual inductance is the same in both directions: a changing current in coil 2 induces an emf $\mathcal{E}_1 = -M\ dI_2/dt$ in coil 1, with the same M.

Self versus mutual. The two effects are the same physics, Faraday's law, seen from different circuits. Self-inductance L describes the emf a coil induces in *itself* when its own current changes; mutual inductance M describes the emf it induces in *another* coil. Every coil has self-inductance; mutual inductance requires a second coil nearby. In a pair of coils, each has its own L, and the pair shares one M. Self-inductance is usually something to be managed, a delay or a spark to be controlled. Mutual inductance is usually the point: it is how energy and signals are passed between circuits that have no wire in common.

Example 25.2: A coil around a solenoid

A long solenoid of length $0.50\ \text{m}$ and cross-sectional area $4.0\ \text{cm}^2$ has 1000 turns. A small coil of 50 turns is wound tightly around its middle. (a) Find the mutual inductance. (b) What emf is induced in the small coil when the solenoid's current changes at $100\ \text{A/s}$?

Solution

(a) The solenoid's field inside is $B_1 = \mu_0(N_1/\ell)I_1$, and all of it passes through each turn of the small coil, so $\Phi_{21} = B_1A$ and

$$M = \frac{N_2\Phi_{21}}{I_1} = \frac{\mu_0 N_1 N_2 A}{\ell} = \frac{(4\pi\times10^{-7})(1000)(50)(4.0\times10^{-4})}{0.50} = \boxed{5.0\times10^{-5}\,\text{H}}. \tag{25.1.3}$$

(b) $|\mathcal{E}_2| = M\ dI_1/dt = (5.0\times10^{-5})(100) = \boxed{5.0\,\text{mV}}$. The size of the small coil does not matter, as long as it encloses the solenoid, since the solenoid's field outside is negligible.

Note: Mutual inductance at work

A wireless phone charger is a flat coil in the charging pad, driven by a rapidly alternating current, and a second coil inside the phone. The changing flux of the first induces an emf in the second, which charges the battery. Electric toothbrushes, implanted medical devices such as pacemakers and cochlear implants, and the contactless payment cards held near a reader all receive energy or data the same way. The coupling works only over short distances, because the field of a coil falls off rapidly, as the field of a dipole does (Section 23.3).

Checkpoint 25.1

(a) A solenoid's number of turns is doubled, keeping its length and area the same. By what factor does its inductance change? (b) Two coils are placed side by side. How could you arrange them to make their mutual inductance as small as possible?

Answer

Answers to Checkpoints

Checkpoint 25.1. (a) The inductance of a solenoid, $L = \mu_0 N^2 A/\ell$, goes as N^2: doubling the turns doubles the field each turn produces *and* doubles the number of turns that field threads, so L quadruples. (b) Minimize the flux of one coil that passes through the other: place them with their axes perpendicular, so the field of one runs across, not through, the turns of the other, and move them as far apart as possible.

This page titled 25.1: Inductors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.2: Energy in a Magnetic Field

Building up a current in an inductor takes work, because the source must push charge against the self-induced emf. When the current is I and increasing at the rate dI/dt, the source delivers power $P = |\mathcal{E}_L|I = LI\ dI/dt$ to the inductor. The total work to bring the current from zero to I is

$$U = \int P\,dt = \int_0^I LI'\,dI' = \frac{1}{2}LI^2. \tag{25.2.1}$$

That energy is not lost: it is stored in the inductor, and it is returned when the current decreases.

Key Equation: Energy stored in an inductor

$$U_B = \frac{1}{2}LI^2. \tag{25.6}$$

Compare the energy stored in a capacitor, $U_E = \frac{1}{2}CV^2$ (equation (19.5)), and the kinetic energy of a moving mass, $\frac{1}{2}mv^2$. Inductance plays the role of mass, and current that of velocity.

Energy density. Where is the energy stored? For a capacitor, we found that it resides in the electric field (Section 19.2). For an inductor, it resides in the magnetic field. For a long solenoid, $L = \mu_0 n^2 A\ell$ and $B = \mu_0 nI$, so

$$U_B = \frac{1}{2}\mu_0 n^2 A\ell\, I^2 = \frac{(\mu_0 nI)^2}{2\mu_0}\,A\ell = \frac{B^2}{2\mu_0}\,(A\ell). \tag{25.2.2}$$

The field fills the volume $A\ell$ inside the solenoid, and nearly vanishes outside, so the energy per unit volume, the **energy density** of the magnetic field, is

Key Equation: Energy density of a magnetic field

$$u_B = \frac{B^2}{2\mu_0}. \tag{25.7}$$

Although derived for a solenoid, this holds for any magnetic field in vacuum, just as $u_E = \frac{1}{2}\epsilon_0 E^2$ (equation (19.6)) holds for any electric field. Where both fields are present, the total energy density is $u = \frac{1}{2}\epsilon_0 E^2 + B^2/2\mu_0$. In an electromagnetic wave the two terms turn out to be exactly equal (Chapter 27).

Example 25.3: Electric and magnetic energy near the ground

Near the ground on a fair day, Earth's electric field is about $150\ \mathrm{V/m}$ and its magnetic field about $50\ \mu\mathrm{T}$. Compare their energy densities.

Solution

$$u_E = \frac{1}{2}\epsilon_0 E^2 = \frac{1}{2}(8.85\times10^{-12})(150)^2 = 1.0\times10^{-7}\,\mathrm{J/m^3}, \tag{25.2.3}$$

$$u_B = \frac{B^2}{2\mu_0} = \frac{(5.0\times10^{-5})^2}{2(4\pi\times10^{-7})} = 9.9\times10^{-4}\,\mathrm{J/m^3}. \tag{25.2.4}$$

The magnetic energy density is $\boxed{\text{about ten thousand times}}$ larger. Fields that seem comparable in everyday terms can store very different amounts of energy. A field of $1\ \mathrm{T}$ stores $u_B = 4.0\times10^5\ \mathrm{J/m^3}$; to store as much in an electric field would take $E = 3.0\times10^8\ \mathrm{V/m}$, a hundred times more than air can withstand. This is why large amounts of energy are much more easily stored in magnetic fields than in electric fields.

Example 25.4: The energy in an MRI magnet

The bore of an MRI magnet is a cylinder of radius $0.30\ \text{m}$ and length $1.5\ \text{m}$, with a uniform field of $1.5\ \text{T}$. Estimate the energy stored in the field inside the bore.

Solution

The energy density is $u_B = (1.5)^2/2(4\pi \times 10^{-7}) = 9.0 \times 10^5\ \text{J/m}^3$, and the volume is $\pi(0.30)^2(1.5) = 0.42\ \text{m}^3$, so

$$U = u_B V = (9.0 \times 10^5)(0.42) = \boxed{3.8 \times 10^5\ \text{J}}. \tag{25.2.5}$$

Including the field in the windings and around the magnet, the total is several megajoules, about the kinetic energy of a loaded truck at highway speed. If a superconducting magnet suddenly loses its superconductivity, an event called a *quench*, all of this energy turns to heat within seconds, boiling off the liquid helium. Magnets are designed to quench safely, venting the helium outside the building.

Checkpoint 25.2

(a) The current in an inductor is doubled. By what factor does its stored energy change? (b) In a certain region the magnetic field is tripled. By what factor does the energy density there change?

Answer

Answers to Checkpoints

Checkpoint 25.2. (a) $U_B = \frac{1}{2}LI^2$, so doubling the current quadruples the energy. (b) The energy density is $u_B = B^2/2\mu_0$, so tripling the field multiplies it by 9. In both cases the energy grows as the *square* of the quantity, exactly as for a capacitor's energy and its voltage or electric field.

This page titled 25.2: Energy in a Magnetic Field is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.3: RL Circuits

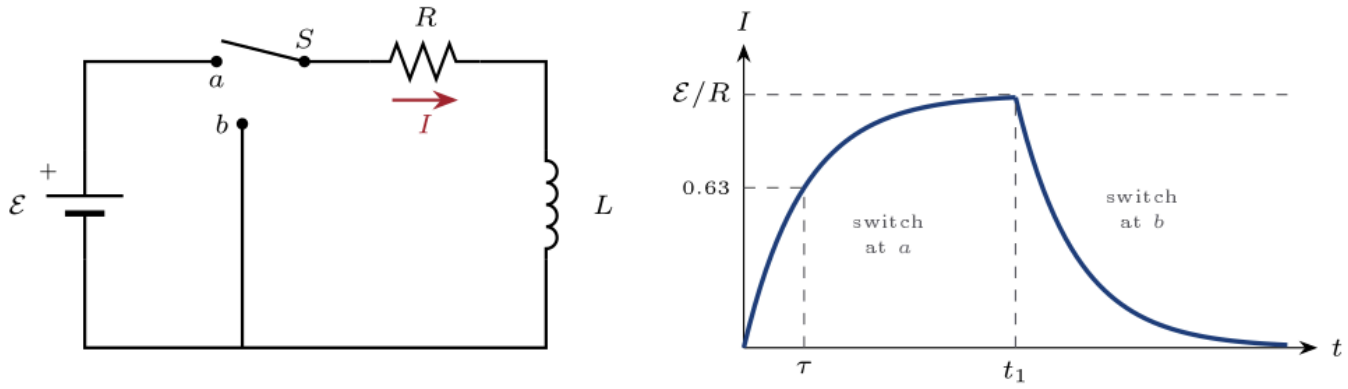


Figure 25.3. An RL circuit. With the switch at a, the current rises toward $\mathcal{E}/R$ with time constant $\tau = L/R$; moved to b at time t_1, the switch removes the battery, and the current decays with the same time constant.

•Figure description

Left: an RL circuit. A battery of emf script E, positive terminal up, is on the left. Its top wire runs to contact a of a two-position switch S, whose pivot connects through a resistor R in the top wire to an inductor L on the right side, which returns along the bottom wire to the battery. A second contact b, the same distance from the pivot as a but below and to the left of it, connects straight down to the bottom wire, so that in position b the resistor and inductor form a loop without the battery. Right: a graph of the current against time. With the switch at a, starting at t equals zero, the current rises from zero toward script E over R, reaching 63 percent of that value at one time constant, tau equals L over R. At a later time t sub 1 the switch is moved to b, and the current decays exponentially back toward zero with the same time constant.

A circuit containing a resistor and an inductor, an **RL circuit**, behaves much like the RC circuit of Section 21.5, but with the roles of current and charge exchanged. In Figure 25.3, the switch can connect a battery to the resistor and inductor in series (position a), or remove the battery and connect the resistor and inductor to each other (position b).

Current rising. Throw the switch to a at $t = 0$. Going around the loop in the direction of the current, the loop rule gives

$$\mathcal{E} - IR - L\frac{dI}{dt} = 0. \tag{25.8}$$

At the first instant, the current is zero, and the whole emf appears across the inductor: $dI/dt = \mathcal{E}/L$. The inductor behaves like a break in the circuit, the opposite of an uncharged capacitor. As the current grows, more of the emf is taken up by the resistor, and the current rises more slowly, approaching the final value $I_f = \mathcal{E}/R$ at which $dI/dt = 0$ and the inductor behaves like a plain wire. Equation (25.8) has the same form as the equation for a charging capacitor, with L in place of R, R in place of $1/C$, and I in place of q, and it has the same kind of solution:

Key Equation: Current rising in an RL circuit

$$I(t) = \frac{\mathcal{E}}{R}\left(1 - e^{-t/\tau}\right), \qquad \tau = \frac{L}{R}. \tag{25.9}$$

The **inductive time constant** $\tau = L/R$ has units of seconds: a henry is a $\mathrm{V \cdot s/A}$ and an ohm is a $\mathrm{V/A}$, so $1\ \mathrm{H}/1\ \Omega = 1\ \mathrm{s}$. After one time constant the current has reached 63% of its final value. A larger inductance slows the rise, since the inductor opposes change more strongly; a larger resistance speeds it up, because the final current, and so the change the inductor must oppose, is smaller. The emf across the inductor, $L\ dI/dt = \mathcal{E}e^{-t/\tau}$, decays from $\mathcal{E}$ to zero.

Current decaying. Once the current has reached I_0, throw the switch to b. The battery is removed, but the inductor keeps the current flowing, now through the resistor alone, and the loop rule gives $-IR - L\ dI/dt = 0$. The current decays exponentially with the same time constant:

Key Equation: Current decaying in an RL circuit

$$I(t) = I_0\, e^{-t/\tau}, \qquad \tau = \frac{L}{R}. \tag{25.10}$$

During the decay, the energy $\frac{1}{2}LI_0^2$ stored in the inductor's field is delivered to the resistor as heat (Problem 25.18).

Example 25.5: Switching on an RL circuit

A coil with inductance $0.50\ \text{H}$ and resistance $10\ \Omega$ is connected to a $12\ \text{V}$ battery. Find (a) the time constant, (b) the final current, (c) the initial rate of increase of the current, and (d) the time for the current to reach 90% of its final value.

Solution

(a) $\tau = L/R = 0.50/10 = \boxed{0.050\,\text{s}}$.

(b) $I_f = \mathcal{E}/R = 12/10 = \boxed{1.2\,\text{A}}$.

(c) At $t = 0$, the current is zero, so the whole emf is across the inductor: $dI/dt = \mathcal{E}/L = 12/0.50 = \boxed{24\,\text{A/s}}$. If the current kept rising at this rate, it would reach I_f in exactly one time constant.

(d) Setting $1 - e^{-t/\tau} = 0.90$ gives $e^{-t/\tau} = 0.10$, so

$$t = \tau \ln 10 = (0.050)(2.30) = \boxed{0.12\,\text{s}}. \tag{25.3.1}$$

Breaking an inductive circuit. What if the current in an inductor is simply switched off, by opening a switch, with no path like b for it to continue? Then dI/dt is enormous, and so is the self-induced emf, $L\ dI/dt$. It can reach thousands of volts, enough to ionize the air across the opening switch contacts and drive a spark, through which the current, and the energy stored in the field, find a way to escape. This is the spark seen when a plug is pulled from a running motor, and it slowly erodes the contacts of switches and relays. Circuits that switch inductive loads include a path, often a diode, to let the current decay gently instead. The same effect is put to use in a car's ignition coil, which interrupts the current in a coil to produce the tens of kilovolts that fire the spark plugs.

Example 25.6: A spark from an inductor

A relay coil with an inductance of $2.0\ \text{H}$ carries $1.5\ \text{A}$. The circuit is broken, and the current falls to zero in about $1.0\ \text{ms}$. Estimate the induced emf and the energy that must be dissipated.

Solution

$$|\mathcal{E}_L| \approx L\frac{\Delta I}{\Delta t} = (2.0)\frac{1.5}{1.0\times 10^{-3}} = \boxed{3000\,\text{V}}, \tag{25.3.2}$$

far more than the few volts that were driving the coil. The energy to be dissipated is the energy stored in the field, $\frac{1}{2}LI^2 = \frac{1}{2}(2.0)(1.5)^2 = \boxed{2.3\,\text{J}}$, released in a millisecond: a power of more than two kilowatts, concentrated in a tiny spark.

Checkpoint 25.3

In the circuit of Figure 25.3, with the switch at a: (a) What is the potential difference across the inductor just after the switch is closed, and long afterward? (b) Compare with a capacitor in place of the inductor. (c) What is the energy stored in the inductor long after the switch is closed?

Answer

Answers to Checkpoints

Checkpoint 25.3. (a) Just after the switch closes the current is still zero, so there is no voltage across the resistor and the inductor's back emf takes up the entire battery voltage: $V_L = \mathcal{E}$. Long afterward the current is steady, so $dI/dt = 0$ and $V_L = 0$; the inductor behaves like a plain wire. (b) A capacitor behaves the opposite way: just after closing, an uncharged capacitor has $V_C = 0$ and acts like a wire; long afterward it is fully charged, with $V_C = \mathcal{E}$, and carries no current. (c) Long afterward the current is $\mathcal{E}/R$, so $U_B = \frac{1}{2}L(\mathcal{E}/R)^2$.

This page titled 25.3: RL Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.4: Transformers

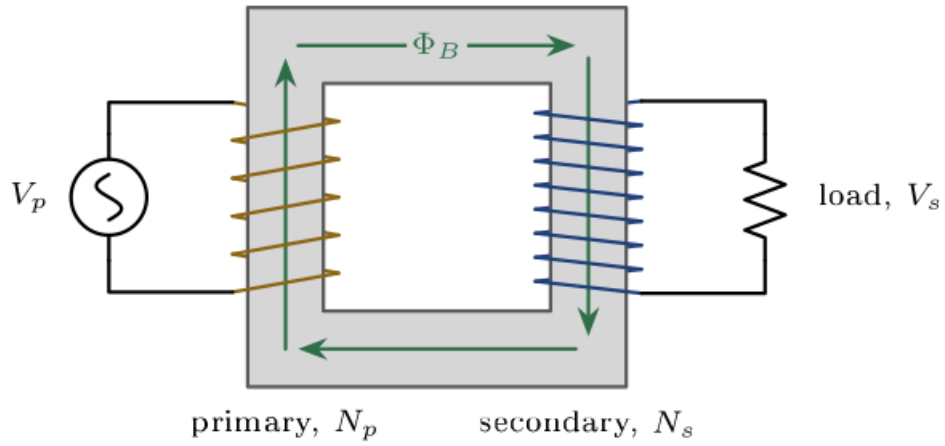


Figure 25.4. A transformer. The alternating current in the primary coil produces an alternating flux that the iron core carries through the secondary coil. The voltages are in the ratio of the numbers of turns.

•Figure description

A transformer. A rectangular iron core, drawn as a thick square ring, has a primary coil of N sub p turns wound on its left side and a secondary coil of N sub s turns, with more turns, wound on its right side. The primary is connected to an alternating voltage source V sub p. The secondary is connected to a load resistor, with voltage V sub s across it. Arrows inside the core show the magnetic flux circulating around the ring, passing through both coils.

Chapter 20 showed why electrical power is transmitted at high voltage: for a given power, a higher voltage means a smaller current, and the power lost in the lines, I^2R, falls as the square of the current (Example 20.10). But generators produce tens of kilovolts, not hundreds, and homes use a hundred or two hundred volts, not hundreds of thousands. The device that changes one voltage into another is the **transformer**. It depends on mutual inductance.

A transformer consists of two coils wound on a common iron core (Figure 25.4). An alternating voltage V_p is applied to the *primary* coil, of N_p turns. The alternating current in it produces an alternating flux in the core, which, because iron has such a large permeability, is confined almost entirely within the core and passes through the *secondary* coil, of N_s turns. For an ideal transformer, with no flux leakage and no losses, the flux Φ_B through every turn of both coils is the same. By Faraday's law, the emf induced in each coil is $N\ d\Phi_B/dt$, and with the same $d\Phi_B/dt$ for both,

Key Equation: The ideal transformer

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}, \qquad \frac{I_s}{I_p} = \frac{N_p}{N_s}. \tag{25.11}$$

The voltage ratio equals the turns ratio. A transformer with more turns on the secondary than the primary is a *step-up* transformer; one with fewer is a *step-down* transformer. The current relation follows from energy conservation: an ideal transformer has no losses, so the power delivered to the secondary circuit equals the power drawn by the primary, $V_pI_p = V_sI_s$. Stepping the voltage up steps the current down in the same ratio. A transformer does not create energy; it trades voltage for current.

Example 25.7: A phone charger

An old-style phone charger contains a transformer that steps $120\ \text{V}$ ac down to $6.0\ \text{V}$. (a) If the primary has 1200 turns, how many does the secondary have? (b) If the phone draws $1.0\ \text{A}$ from the secondary, what current does the primary draw, assuming an ideal transformer?

Solution

(a) $N_s = N_p\ (V_s/V_p) = (1200)(6.0/120) = \boxed{60}$ turns.

(b) $I_p = I_s\ (N_s/N_p) = (1.0)(60/1200) = \boxed{0.050\ \text{A}}$. The power is $6.0\ \text{W}$ on both sides. (Modern chargers use electronic switching circuits and much smaller, higher-frequency transformers, but the principle is the same.)

Example 25.8: From power station to home

A generator produces power at $20\ \text{kV}$. A step-up transformer raises this to $400\ \text{kV}$ for transmission, and near the city a chain of step-down transformers lowers it to $240\ \text{V}$ for homes. (a) What is the turns ratio of the step-up transformer? (b) By what factor does stepping up the voltage reduce the power lost in the transmission lines, compared with transmitting at $20\ \text{kV}$?

Solution

(a) $N_s/N_p = V_s/V_p = 400/20 = \boxed{20}$.

(b) For the same power, the current in the lines is reduced by the factor 20. The loss I^2R is reduced by $20^2 = \boxed{400}$. Without transformers, alternating current would have no advantage over direct current, and a national grid would waste most of its power as heat. That is why George Westinghouse's ac system, using transformers, won out over Thomas Edison's dc system in the "war of the currents" of the 1880s and 1890s. (Today some very long lines use direct current at high voltage, with electronic converters at each end, but the power still passes through transformers on the way.)

Only for changing currents. A transformer works only if the flux in its core changes. Connect its primary to a battery, and after the brief transient of Section 25.3 the current is steady, the flux is constant, and nothing is induced in the secondary. That is the main reason why electrical power is distributed as alternating current (Chapter 26).

Real transformers. Large transformers are more than 99% efficient, but not perfectly so. Their losses come from the resistance of the windings; from eddy currents in the core, reduced by building the core from thin insulated sheets of iron (Section 24.2); and from hysteresis in the core as it is magnetized back and forth, reduced by using soft magnetic material (Section 23.7). The energy lost appears as heat, which is why large transformers are cooled by circulating oil, and as the familiar hum, as the core vibrates at twice the frequency of the ac supply.

Checkpoint 25.4

(a) A transformer's secondary has half as many turns as its primary. Is it a step-up or a step-down transformer? How do the secondary voltage and current compare with the primary's? (b) Why would connecting a transformer's primary to a car battery accomplish nothing, except perhaps to overheat the primary coil?

Answer

Answers to Checkpoints

Checkpoint 25.4. (a) Step-down. For an ideal transformer, $V_s/V_p = N_s/N_p = \frac{1}{2}$, so the secondary voltage is half the primary's. Power in equals power out, $I_pV_p = I_sV_s$, so the secondary current is *twice* the primary's. (b) A transformer works by induction, which requires a *changing* flux. A battery's steady current makes a steady flux, so after an instant no emf is induced in the secondary: there is no output. Worse, with no changing flux there is no back emf in the primary to limit its current, only the coil's small resistance, so a large steady current flows and heats the primary.

This page titled 25.4: Transformers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.S: Inductance (Summary)

Summary

- A coil's changing current induces an emf in the coil itself, $\mathcal{E}_L = -L\ dI/dt$, where the **self-inductance** $L = N\Phi_B/I$ depends on geometry. The unit is the henry. For a long solenoid, $L = \mu_0 N^2 A/\ell$. An inductor opposes changes in current, not current itself.
- Two coils are coupled by their **mutual inductance** M: a changing current in one induces $\mathcal{E}_2 = -M\ dI_1/dt$ in the other, and M is the same in both directions.
- An inductor stores energy $U_B = \frac{1}{2}LI^2$ in its magnetic field. Any magnetic field has energy density $u_B = B^2/2\mu_0$, the counterpart of $u_E = \frac{1}{2}\epsilon_0 E^2$.
- In an **RL circuit**, the current rises as $I = (\mathcal{E}/R)(1 - e^{-t/\tau})$ and decays as $I = I_0 e^{-t/\tau}$, with time constant $\tau = L/R$. Interrupting the current in an inductor suddenly produces a large emf and a spark.
- A **transformer** uses mutual inductance to change an alternating voltage: $V_s/V_p = N_s/N_p$ and, for an ideal transformer, $I_s/I_p = N_p/N_s$. Stepping up the voltage for transmission cuts the I^2R losses in the lines.

Key Equations

Self-inductance

$$L = \frac{N\Phi_B}{I}, \qquad \mathcal{E}_L = -L\frac{dI}{dt}, \qquad L_{\text{solenoid}} = \frac{\mu_0 N^2 A}{\ell} \tag{25.S.1}$$

Mutual inductance

$$M = \frac{N_2\Phi_{21}}{I_1}, \qquad \mathcal{E}_2 = -M\frac{dI_1}{dt} \tag{25.S.2}$$

Magnetic energy

$$U_B = \frac{1}{2}LI^2, \qquad u_B = \frac{B^2}{2\mu_0} \tag{25.S.3}$$

RL circuits

$$I = \frac{\mathcal{E}}{R}\left(1 - e^{-t/\tau}\right), \qquad I = I_0 e^{-t/\tau}, \qquad \tau = \frac{L}{R} \tag{25.S.4}$$

Ideal transformer

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}, \qquad \frac{I_s}{I_p} = \frac{N_p}{N_s} \tag{25.S.5}$$

This page titled 25.S: Inductance (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

25.E: Inductance (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T\cdot m/A}$.

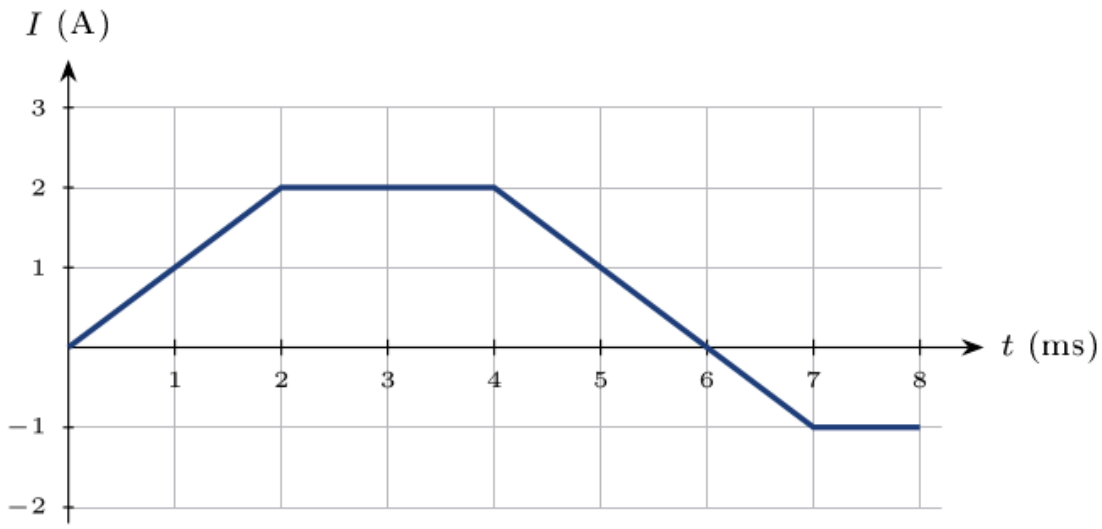
Section 25.1 — Inductors

25.1 ★ The current in a $25\ \mathrm{mH}$ inductor changes from $2.0\ \mathrm{A}$ to $5.0\ \mathrm{A}$ in $0.010\ \mathrm{s}$. Find the magnitude of the self-induced emf.

25.2 ★ A current of $3.0\ \mathrm{A}$ in a coil of 200 turns produces a flux of $1.5 \times 10^{-4}\ \mathrm{Wb}$ through each turn. Find the coil's inductance.

25.3 ★★ What length of wire is needed to wind a $1.0\ \mathrm{mH}$ air-core solenoid $10\ \mathrm{cm}$ long and $2.0\ \mathrm{cm}$ in diameter?

25.4 ★★ *[Graph-reading.]* The graph shows the current in a $40\ \mathrm{mH}$ inductor as a function of time.



•Figure description
A graph of current in amperes, from minus 2 to 3, against time in milliseconds, from 0 to 8. The current rises linearly from 0 at t equals 0 to 2 amperes at 2 milliseconds, stays at 2 amperes until 4 milliseconds, then falls linearly to minus 1 ampere at 7 milliseconds, and stays there until 8 milliseconds.

a. Find the emf across the inductor in each time interval.
b. Sketch the emf as a function of time.
c. At what time is the energy stored in the inductor largest? Is it ever zero after $t = 0$?

25.5 ★★ *[Conceptual.]* Two inductors are made of the same length of the same wire. One is wound as a long, thin solenoid and the other as a short, fat one. Which has the larger inductance? (Use equation (25.4), keeping the total length of wire fixed.)

25.6 ★★★ A toroid has N turns, inner radius a, outer radius b, and a rectangular cross-section of height h. Using $B = \mu_0 N I/2\pi r$ (equation (23.10)), show that its inductance is $L = (\mu_0 N^2 h/2\pi)\ln(b/a)$.

Section 25.1.1 — Mutual vs. Self-Inductance

25.7 ★ Two coils have a mutual inductance of $3.0\ \mathrm{mH}$. The current in the first changes at $800\ \mathrm{A/s}$. What emf is induced in the second?

25.8 ★★ A short coil of 20 turns and radius $1.0\ \mathrm{cm}$ lies at the center of a large circular coil of 100 turns and radius $20\ \mathrm{cm}$, in the same plane. Find their mutual inductance, treating the field of the large coil as uniform over the small one.

25.9 ★★ *[Conceptual.]* Two identical coils are placed (a) side by side with their axes parallel, (b) coaxially, one in front of the other, and (c) with their axes perpendicular. Rank the arrangements by mutual inductance.

25.10 ★★★ Two coils have self-inductances L_1 and L_2 and are wound tightly together, so that all of the flux of each passes through every turn of the other. Show that their mutual inductance is $M = \sqrt{L_1 L_2}$.

Section 25.2 — Energy in a Magnetic Field

25.11 ★ How much energy is stored in a $50\ \text{mH}$ inductor carrying $4.0\ \text{A}$?

25.12 ★★ Find the energy density of the magnetic field at a distance of $1.0\ \text{cm}$ from a long straight wire carrying $20\ \text{A}$.

25.13 ★★ *[Real-world: energy storage.]* Superconducting magnetic energy storage systems store energy in the field of a large superconducting coil. What field would be needed to store $100\ \text{MJ}$ in a volume of $50\ \text{m}^3$? What current would a $20\ \text{H}$ coil need to carry to store this much?

25.14 ★★★ Use the energy density u_B to find the energy stored per unit length in the magnetic field *inside* a long straight wire of radius R carrying a uniformly distributed current I. Show that it does not depend on R.

Section 25.3 — RL Circuits

25.15 ★ Find the time constant of an RL circuit with $L = 0.20\ \text{H}$ and $R = 40\ \Omega$. How long does the current take to reach half its final value?

25.16 ★★ A $9.0\ \text{V}$ battery is connected to a $150\ \text{mH}$ inductor in series with a $30\ \Omega$ resistor. Find the current and the rate of change of current (a) just after the connection, (b) after one time constant, and (c) long afterward.

25.17 ★★ *[Real-world: measurement.]* In an RL circuit connected to a $6.0\ \text{V}$ battery, the current is measured as it rises. It reaches $0.19\ \text{A}$ after $2.0\ \text{ms}$ and levels off at $0.30\ \text{A}$. Find R, the time constant, and L.

25.18 ★★ Show, by integrating I^2R over time, that the energy dissipated in the resistor while the current decays in an RL circuit equals $\tfrac{1}{2}LI_0^2$, the energy initially stored in the inductor.

25.19 ★★★ In an RL circuit with the current rising, find the time at which the rate at which energy is being stored in the inductor is greatest. Express your answer in terms of τ.

Section 25.4 — Transformers

25.20 ★ A transformer has 500 turns on its primary and 25 on its secondary. If the primary is connected to $120\ \text{V}$ ac, what is the secondary voltage?

25.21 ★★ A neon sign needs $12\ \text{kV}$ and draws $30\ \text{mA}$. It is supplied from $120\ \text{V}$ through a transformer. Find the turns ratio and the primary current, assuming an ideal transformer.

25.22 ★★ *[Real-world: the grid.]* A town uses an average of $50\ \text{MW}$, delivered through lines with a total resistance of $0.40\ \Omega$. Find the power lost in the lines if the power is transmitted at (a) $20\ \text{kV}$ and (b) $230\ \text{kV}$, and the fraction of the delivered power lost in each case.

25.23 ★★ *[Conceptual.]* A transformer's secondary is left open, with nothing connected. Explain why the primary then draws very little current, even though it is connected to the full supply voltage. (Hint: think about the inductance of the primary coil.)

25.24 ★★★ A resistor R is connected to the secondary of an ideal transformer with turns ratio N_p/N_s. Show that, as seen from the primary, it behaves like a resistor $R' = (N_p/N_s)^2 R$. (This "impedance matching" is how transformers are used to deliver the most power to a loudspeaker; compare Section 21.4.)

This page titled 25.E: Inductance (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26: AC CIRCUITS

The electricity delivered to your home does not flow steadily in one direction. Sixty times a second in North America, fifty in most of the world, the voltage at the outlet swings from positive to negative and back, and the current in every lamp and appliance reverses direction with it. This is alternating current, and the whole electrical grid is built on it. Part of the reason is that generators produce ac naturally (Section 24.4). The larger reason is the long trip from the power plant to your home. Power travels efficiently only at high voltage, where the current, and with it the I^2R loss in the lines, is small (Example 20.10). Nobody wants hundreds of kilovolts in a kitchen, though, so the voltage must be stepped up for the journey and back down at the end, and as Chapter 25 showed, transformers can do that only for a changing current. Ac is also easier to switch off safely: the current passes through zero twice in every cycle, which helps a circuit breaker interrupt it cleanly. Today, power electronics can convert between ac and dc at high voltage, and some of the longest transmission lines and undersea cables now carry dc, but the grid that reaches your outlets is still ac.

Alternating currents make circuit elements behave in new ways. A capacitor, which blocks a steady current completely, passes an alternating one, and passes it more easily the faster it alternates. An inductor does the opposite. Combine them, and a circuit can respond strongly to one frequency and ignore all others, which is how a radio picks out a single station from the hundreds of signals reaching its antenna at once. This chapter begins with the natural oscillations of a circuit containing only a capacitor and an inductor, then follows resistors, capacitors, and inductors driven by an alternating source, one at a time and together, and ends with the power that ac circuits deliver.

Learning Objectives

- Describe the oscillation of charge and energy in an LC circuit, and find its frequency.
- Use rms values and phasors to describe alternating currents and voltages.
- Find the current in a resistor, a capacitor, and an inductor driven by an ac source, using reactance and phase.
- Analyze a series RLC circuit, using impedance and the phase constant, and find its resonant frequency.
- Calculate the average power delivered to an ac circuit, using the power factor.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

26.1: LC Oscillations
26.2: Resistors in an AC Circuit
26.3: Capacitors in an AC Circuit
26.4: Inductors in an AC Circuit
26.5: Series RLC Circuits
26.6: Power in AC Circuits
26.S: AC Circuits (Summary)
26.E: AC Circuits (Problems)

This page titled 26: AC Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.1: LC Oscillations

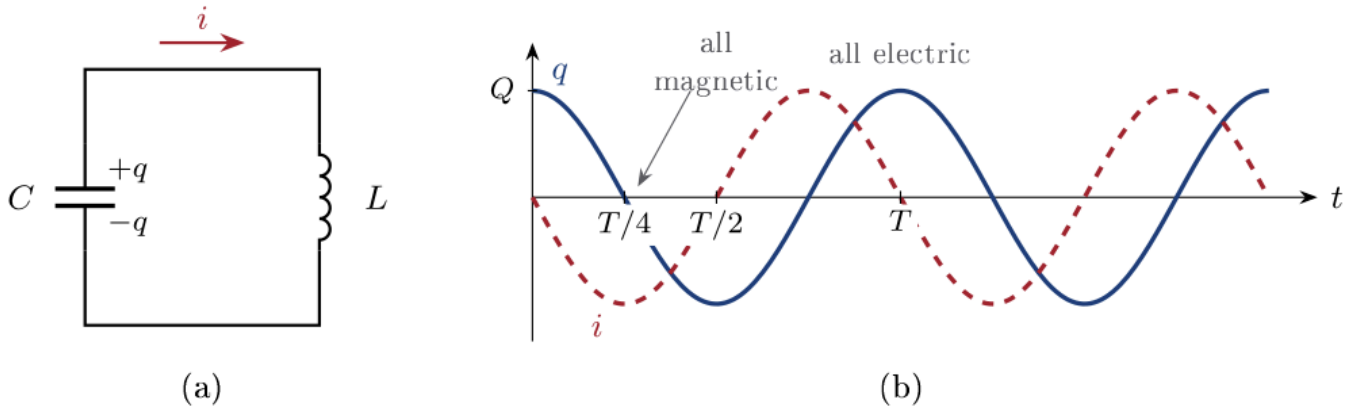


Figure 26.1. (a) An LC circuit. (b) The charge on the capacitor (solid) and the current (dashed) oscillate sinusoidally, a quarter cycle apart. When the charge is greatest, the current is zero and all the energy is in the capacitor; a quarter period later, the charge is zero, the current is greatest, and all the energy is in the inductor.

•Figure description

Panel a: an LC circuit, a capacitor C on the left connected in a loop with an inductor L on the right. The capacitor's upper plate carries charge q, and current i flows around the loop. Panel b: a graph against time of the charge q, a cosine curve starting at its maximum Q, and the current i, a negative sine curve that is a quarter cycle behind it. Labels mark that when the charge is at a maximum the energy is all electric, in the capacitor, and when the current is at a maximum a quarter period later the energy is all magnetic, in the inductor.

Charge a capacitor, then connect it across an inductor (Figure 26.1a). The capacitor begins to discharge through the inductor, but the inductor opposes any change in current (Section 25.1), so the current builds up gradually. By the time the capacitor is fully discharged, the current is at its largest, and all the energy that was stored in the capacitor's electric field is now in the inductor's magnetic field. The inductor keeps the current flowing, and the capacitor charges up again, with the opposite polarity. Then the process reverses. With no resistance to dissipate it, the energy sloshes back and forth between the two fields forever, and the charge and current oscillate. This is an **LC circuit**.

The equation of motion. Let q be the charge on the capacitor's upper plate and $i = dq/dt$ the current. Going around the loop, the potential changes by $-q/C$ across the capacitor and by $-L\ di/dt$ across the inductor, and the loop rule gives

$$-\frac{q}{C} - L\frac{di}{dt} = 0 \qquad \Longrightarrow \qquad \frac{d^2q}{dt^2} = -\frac{1}{LC}\,q. \tag{26.1.1}$$

The second derivative of q is proportional to q itself, with a negative constant. The functions whose second derivatives are proportional to their negatives are sines and cosines. Try $q = Q\cos(\omega t + \phi)$: then $d^2q/dt^2 = -\omega^2 Q\cos(\omega t + \phi) = -\omega^2 q$, which satisfies the equation if $\omega^2 = 1/LC$. So

Key Equation: LC oscillations

$$q = Q\cos(\omega t + \phi), \qquad i = -\omega Q\sin(\omega t + \phi), \qquad \omega = \frac{1}{\sqrt{LC}}. \tag{26.1}$$

The charge oscillates sinusoidally with angular frequency $\omega = 1/\sqrt{LC}$, and frequency $f = \omega/2\pi$. The amplitude Q and the phase constant ϕ are set by how the oscillation starts; if the capacitor is fully charged at $t = 0$, then $\phi = 0$. The current oscillates with the same frequency, a quarter cycle behind the charge, and its amplitude is $I = \omega Q$ (Figure 26.1b).

Energy. The energy in the capacitor is $U_E = q^2/2C = (Q^2/2C)\cos^2(\omega t + \phi)$, and the energy in the inductor is $U_B = \frac{1}{2}Li^2 = \frac{1}{2}L\omega^2Q^2\sin^2(\omega t + \phi) = (Q^2/2C)\sin^2(\omega t + \phi)$, using $\omega^2 = 1/LC$. Their sum is

$$U = U_E + U_B = \frac{Q^2}{2C}\left[\cos^2(\omega t + \phi) + \sin^2(\omega t + \phi)\right], \tag{26.1.2}$$

and since $\cos^2 + \sin^2 = 1$,

Key Equation: Energy in an LC circuit

$$U = \frac{q^2}{2C} + \frac{1}{2}Li^2 = \frac{Q^2}{2C} = \frac{1}{2}LI^2, \tag{26.2}$$

which is constant, as it must be with no resistance. At every instant the energy is shared between the two fields, but the total is always the energy the capacitor started with, $Q^2/2C$, or equally the energy the inductor holds when the current peaks, $\frac{1}{2}LI^2$. The energy passes back and forth between the electric and magnetic fields twice in each cycle.

A mechanical twin. The LC equation has exactly the form of the equation of motion of a mass on a spring, $d^2x/dt^2 = -(k/m)x$, with charge in place of position, current in place of velocity, L in place of the mass m, and $1/C$ in place of the spring constant k. The capacitor's energy $q^2/2C$ corresponds to the spring's potential energy $\frac{1}{2}kx^2$, and the inductor's $\frac{1}{2}Li^2$ to the kinetic energy $\frac{1}{2}mv^2$. Chapter 28 develops this kind of motion, simple harmonic motion, in detail. A real circuit always has some resistance, which dissipates energy and makes the oscillations die away, just as friction damps a mechanical oscillator (Section 28.8).

Example 26.1: An LC circuit

A $10\ \mu\text{F}$ capacitor charged to $20\ \text{V}$ is connected to a $25\ \text{mH}$ inductor at $t = 0$. Find (a) the angular frequency and frequency of the oscillations, (b) the maximum charge and maximum current, and (c) the total energy.

Solution

(a) The angular frequency is

$$\omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{(0.025)(10\times10^{-6})}} = \boxed{2000\,\text{rad/s}}, \tag{26.1.3}$$

and $f = \omega/2\pi = \boxed{320\,\text{Hz}}$.

(b) $Q = CV = (10\times10^{-6})(20) = \boxed{200\,\mu\text{C}}$, and $I = \omega Q = (2000)(200\times10^{-6}) = \boxed{0.40\,\text{A}}$.

(c) The energy is all in the capacitor when its charge is greatest:

$$U = \frac{Q^2}{2C} = \frac{1}{2}CV^2 = \frac{1}{2}(10\times10^{-6})(20)^2 = \boxed{2.0\,\text{mJ}}. \tag{26.1.4}$$

As a check, the energy in the inductor when the current is greatest is $\frac{1}{2}LI^2 = \frac{1}{2}(0.025)(0.40)^2 = 2.0\ \text{mJ}$, the same.

Example 26.2: Tuning a radio

The tuning circuit of an AM radio has a $0.20\ \text{mH}$ inductor and a variable capacitor. What range of capacitance lets it tune across the AM band, from $540\ \text{kHz}$ to $1600\ \text{kHz}$?

Solution

Solving $f = 1/(2\pi\sqrt{LC})$ for C gives $C = 1/(4\pi^2 f^2 L)$. At $540\ \text{kHz}$,

$$C = \frac{1}{4\pi^2(540\times10^3)^2(0.20\times10^{-3})} = \boxed{430\,\text{pF}}, \tag{26.1.5}$$

and at $1600\ \text{kHz}$, $C = \boxed{49\,\text{pF}}$. Turning the tuning knob changes the overlap of the capacitor's plates between these limits. Because f depends on $1/\sqrt{C}$, a factor of three in frequency needs a factor of nine in capacitance. How the circuit singles out one frequency is the subject of Section 26.5.

Checkpoint 26.1

In an LC circuit oscillating with period T: (a) Where is all of the energy at the instant the capacitor's charge passes through zero? What is the current doing then? (b) How many times during one period is all of the energy stored in the capacitor? (c) Which mechanical oscillator is this like, and what plays the role of the spring and of the mass?

Answer

Answers to Checkpoints

Checkpoint 26.1. (a) All of it is in the inductor's magnetic field. With $q = 0$ the capacitor stores nothing, so the energy $\frac{1}{2}LI^2$ is at its maximum, and the current is at its peak value (momentarily not changing). (b) Twice. The capacitor is fully charged once with one polarity and, half a period later, once with the opposite polarity; the current is zero at both instants. (c) A mass on a spring. The capacitor's charge plays the role of the displacement and the current the role of the velocity: $1/C$ acts like the spring constant, storing energy in the "stretched" capacitor, and L acts like the mass, the inertia that keeps the current flowing.

This page titled 26.1: LC Oscillations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.2: Resistors in an AC Circuit

An ac generator (Section 24.4) produces a sinusoidal emf. We write the voltage of such a source as

$$v = V_m \sin \omega t, \tag{26.2.1}$$

where V_m is its amplitude, and $\omega = 2\pi f$ is the *driving* angular frequency, set by the source. Lowercase letters v and i stand for the instantaneous values, which change with time; capitals stand for amplitudes and other constants. The current that such a source drives is an **alternating current**.

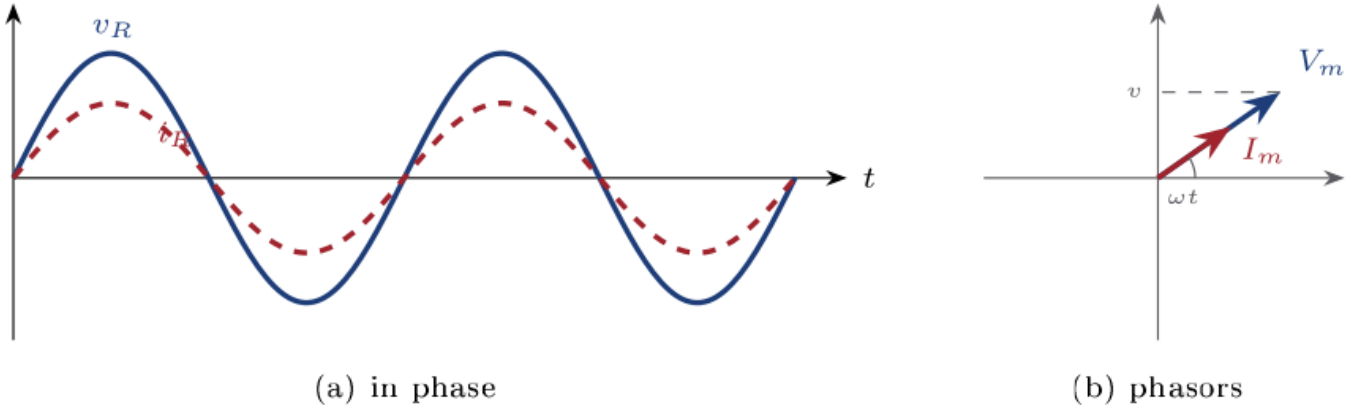


Figure 26.2. A resistor in an ac circuit. (a) The current and voltage are in phase. (b) Their phasors point in the same direction and rotate together; the vertical components are the instantaneous values.

•Figure description

Panel a: graphs against time of the voltage v across a resistor and the current i through it, both sine curves that reach their peaks and zeros at the same times: they are in phase. Panel b: a phasor diagram. The voltage phasor, of length V sub m, and the current phasor, of length I sub m, point in the same direction, at angle omega t above the horizontal axis, and rotate counterclockwise together. Their projections onto the vertical axis are the instantaneous values v and i.

Connect the source across a resistor R. The loop rule gives $v - iR = 0$ at every instant, so

$$i_R = \frac{V_m}{R} \sin \omega t = I_m \sin \omega t, \qquad I_m = \frac{V_m}{R}. \tag{26.3}$$

The current and the voltage rise and fall together: they are *in phase* (Figure 26.2a). Ohm's law holds for the amplitudes, $V_m = I_m R$, and at every instant.

Phasors. Keeping track of the phases of several sinusoids is much easier with a picture. Represent a quantity that varies as $V_m \sin(\omega t + \phi)$ by a vector of length V_m that rotates counterclockwise with angular velocity ω, making the angle $\omega t + \phi$ with the horizontal axis. Its vertical component is $V_m \sin(\omega t + \phi)$, the instantaneous value. Such a rotating vector is called a **phasor**. Several quantities with the same frequency are drawn as phasors rotating together, and the angles between them show their phase differences. For a resistor, the current phasor and the voltage phasor point in the same direction (Figure 26.2b). (Chapter 29 uses phasors in the same way to add waves.) If phasors are new to you, Section C.8 of Appendix C builds them up step by step, with pictures of how a rotating phasor traces out a sine curve, what it means for one sinusoid to lead another, and how adding phasors adds sinusoids.

Rms values. The average of a sinusoidal current over a cycle is zero, which is not a useful measure of its size. What matters for heating is the average of i^2, since the power in a resistor is $i^2 R$. Because the average of $\sin^2 \omega t$ over a cycle is $\frac{1}{2}$, the average power is $\frac{1}{2} I_m^2 R$. A steady current that delivers the same power is the **root-mean-square**, or rms, current: the square root of the mean of the square,

Key Equation: Rms values

$$I_{\text{rms}} = \frac{I_m}{\sqrt{2}}, \qquad V_{\text{rms}} = \frac{V_m}{\sqrt{2}}. \tag{26.4}$$

Ac meters read rms values, and ac voltages and currents are quoted as rms values unless stated otherwise. The "$120\ \text{V}$" of a North American outlet is an rms voltage; its amplitude is $\sqrt{2}(120) = 170\ \text{V}$. With rms values, the formulas for power in a resistor look just as they do for dc: $P_{\text{avg}} = I_{\text{rms}}^2 R = V_{\text{rms}}^2/R = I_{\text{rms}} V_{\text{rms}}$.

Example 26.3: A lamp on ac

A $60\ \text{W}$ lamp is designed for a $120\ \text{V}$ (rms) outlet. Find (a) its resistance, (b) the rms and peak currents, and (c) the peak power.

Solution

(a) $R = V_{\text{rms}}^2/P_{\text{avg}} = (120)^2/60 = \boxed{240\,\Omega}$, the same lamp as in Chapter 20.

(b) $I_{\text{rms}} = V_{\text{rms}}/R = \boxed{0.50\,\text{A}}$, and $I_m = \sqrt{2}\ I_{\text{rms}} = \boxed{0.71\,\text{A}}$.

(c) The instantaneous power is $i^2R = I_m^2 R \sin^2 \omega t$, which peaks at $I_m^2 R = (0.71)^2(240) = \boxed{120\,\text{W}}$, twice the average. It falls to zero twice each cycle, 120 times a second, but the filament is too hot to cool noticeably in so short a time, so the lamp does not visibly flicker.

This page titled 26.2: Resistors in an AC Circuit is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.3: Capacitors in an AC Circuit

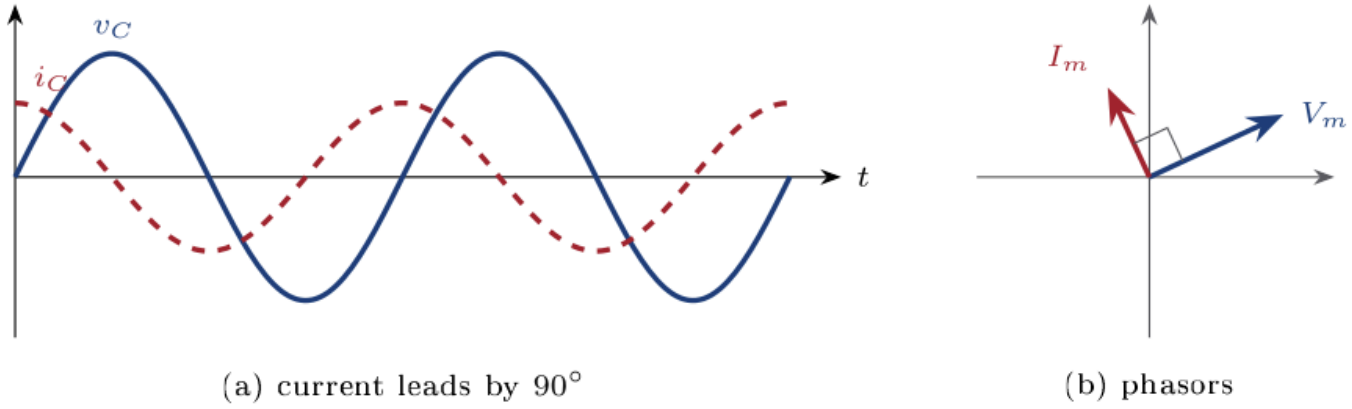


Figure 26.3. A capacitor in an ac circuit. (a) The current reaches its peaks a quarter cycle before the voltage. (b) The current phasor is 90° ahead of the voltage phasor.

•Figure description

Panel a: graphs against time of the voltage v across a capacitor, a sine curve, and the current i, a curve that reaches each peak a quarter cycle earlier than the voltage: the current leads the voltage by 90 degrees. Panel b: phasors. The voltage phasor V sub m is at angle omega t; the current phasor I sub m is 90 degrees ahead of it, counterclockwise.

Now connect the source across a capacitor alone. At every instant the capacitor's voltage equals the source's, so its charge is $q = Cv = CV_m \sin \omega t$, and the current is

$$i_C = \frac{dq}{dt} = \omega C V_m \cos \omega t = \omega C V_m \sin\left(\omega t + 90^\circ\right). \tag{26.3.1}$$

The current in a capacitor *leads* the voltage by 90°, a quarter cycle (Figure 26.3): the current is greatest when the voltage is zero and changing fastest, and zero when the voltage is at its peak and the capacitor is momentarily full. Its amplitude is $I_m = \omega C V_m$. Written like Ohm's law, $V_m = I_m X_C$, with

Key Equation: Capacitive reactance

$$X_C = \frac{1}{\omega C}. \tag{26.5}$$

The quantity X_C is the **capacitive reactance**. It is measured in ohms, and it plays the role of resistance for the amplitudes, although it relates quantities that are out of phase, and, as Section 26.6 shows, a capacitor dissipates no energy. The reactance depends on frequency. At low frequency it is large: at $\omega = 0$, steady dc, a capacitor blocks current completely once it has charged. At high frequency it is small: the capacitor charges and discharges so quickly that it barely builds up any voltage opposing the source. A capacitor passes high frequencies and blocks low ones.

Example 26.4: A capacitor at two frequencies

A $10\ \mu\text{F}$ capacitor is connected to a $120\ \text{V}$ (rms) source. Find the rms current at (a) $60\ \text{Hz}$ and (b) $6.0\ \text{kHz}$.

Solution

(a) The capacitive reactance is

$$X_C = \frac{1}{2\pi f C} = \frac{1}{2\pi(60)(10\times 10^{-6})} = 265\,\Omega, \tag{26.3.2}$$

so $I_{\text{rms}} = V_{\text{rms}}/X_C = 120/265 = \boxed{0.45\ \text{A}}$.

(b) A hundred times the frequency gives a hundredth of the reactance, $X_C = 2.65\ \Omega$, and a current of $\boxed{45\ \text{A}}$. The rms relation $V_{\text{rms}} = I_{\text{rms}}X_C$ works because the rms values are the amplitudes divided by the same $\sqrt{2}$.

This page titled 26.3: Capacitors in an AC Circuit is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.4: Inductors in an AC Circuit

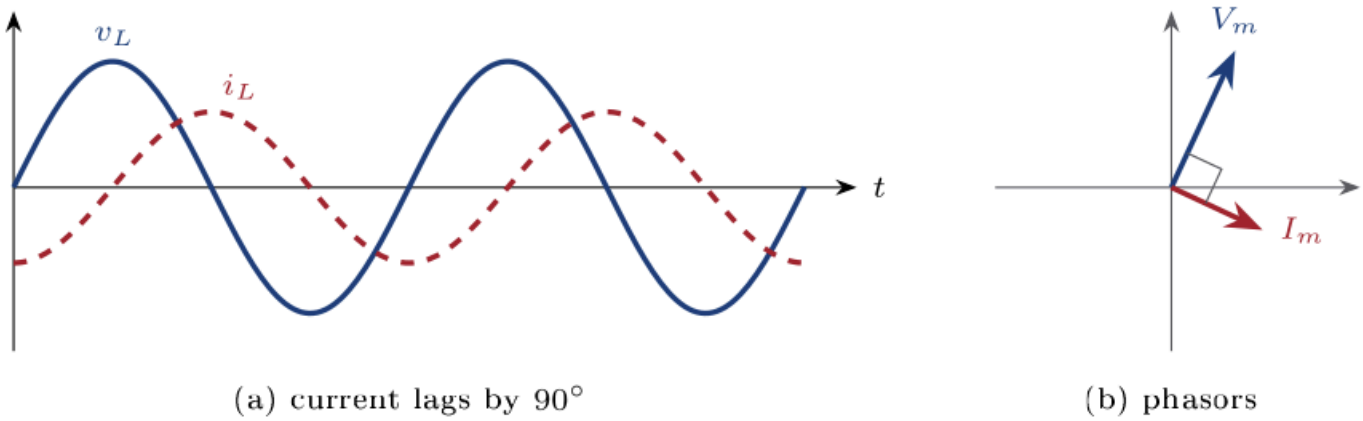


Figure 26.4. An inductor in an ac circuit. (a) The current reaches its peaks a quarter cycle after the voltage. (b) The current phasor is 90° behind the voltage phasor.

•Figure description

Panel a: graphs against time of the voltage v across an inductor, a sine curve, and the current i, a curve that reaches each peak a quarter cycle later than the voltage: the current lags the voltage by 90 degrees. Panel b: phasors. The voltage phasor V sub m is at angle omega t; the current phasor I sub m is 90 degrees behind it, clockwise.

Finally, connect the source across an inductor alone. The loop rule gives $v - L\ di/dt = 0$, so $di/dt = (V_m/L)\sin\omega t$, and integrating,

$$i_L = -\frac{V_m}{\omega L}\cos\omega t = \frac{V_m}{\omega L}\sin\left(\omega t - 90^\circ\right). \tag{26.4.1}$$

The current in an inductor *lags* the voltage by 90° (Figure 26.4). The voltage across an inductor is proportional to the rate of change of the current, so it is greatest when the current is zero and changing fastest. The amplitudes are related by $V_m = I_m X_L$, with

Key Equation: Inductive reactance

$$X_L = \omega L. \tag{26.6}$$

This is the **inductive reactance**, also in ohms. It grows with frequency: at $\omega = 0$ an ideal inductor is just a wire, but at high frequency the rapid changes of current meet a large opposing emf. An inductor passes low frequencies and blocks high ones, the opposite of a capacitor. Loudspeaker systems use this to send the low notes to the large "woofer" through an inductor and the high notes to the small "tweeter" through a capacitor.

A mnemonic for the phases is "ELI the ICE man": in an inductor (L), the emf E comes before the current I; in a capacitor (C), the current I comes before the emf E.

Example 26.5: An inductor at two frequencies

A $0.50\ \text{H}$ inductor with negligible resistance is connected to a $120\ \text{V}$ (rms) source. Find the rms current at (a) $60\ \text{Hz}$ and (b) $6.0\ \text{kHz}$.

Solution

(a) $X_L = 2\pi fL = 2\pi(60)(0.50) = 188\ \Omega$, so $I_{\text{rms}} = 120/188 = \boxed{0.64\ \text{A}}$.

(b) At a hundred times the frequency, $X_L = 18.8\ \text{k}\Omega$, and $I_{\text{rms}} = \boxed{6.4\,\text{mA}}$. Compare Example 26.4: raising the frequency lets a hundred times more current through the capacitor, and a hundred times less through the inductor.

Checkpoint 26.2

(a) What happens to the current through a capacitor, and through an inductor, when the frequency of the source is halved at the same voltage? (b) Which element would you put in series with a loudspeaker to keep low-frequency hum out of it?

Answer

Answers to Checkpoints

Checkpoint 26.2. (a) Halving the frequency doubles the capacitive reactance, $X_C = 1/\omega C$, so the capacitor's current halves. It halves the inductive reactance, $X_L = \omega L$, so the inductor's current doubles. (b) A capacitor in series. Its reactance is large at low frequencies and small at high ones, so it blocks the low-frequency hum while passing the higher-frequency sound. An inductor would do the reverse.

This page titled 26.4: Inductors in an AC Circuit is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.5: Series RLC Circuits

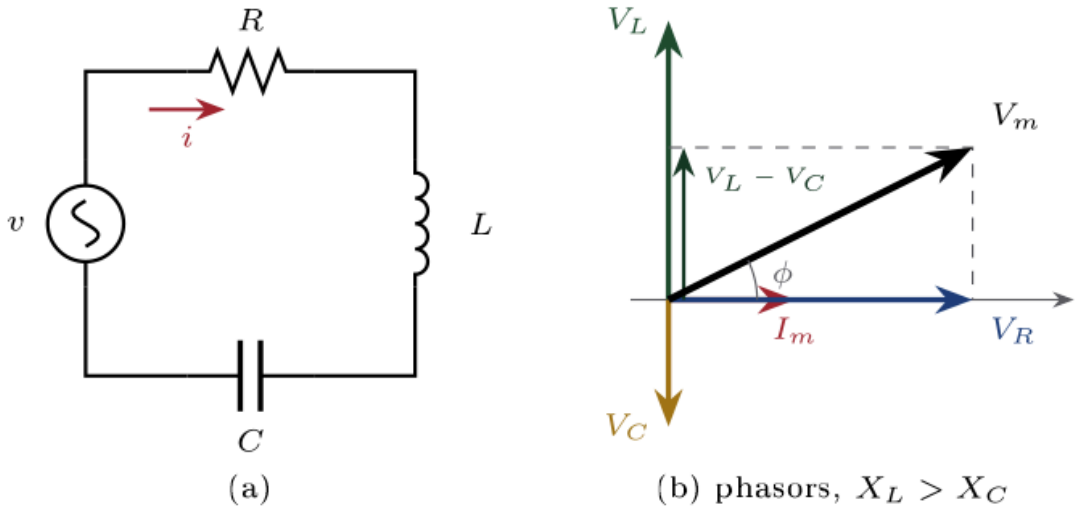


Figure 26.5. (a) A series RLC circuit. (b) Its phasor diagram. The voltage phasors of the resistor, inductor, and capacitor are in phase with, 90° ahead of, and 90° behind the current. They add as vectors to the source voltage V_m, which leads the current by ϕ.

•Figure description

Panel a: a series RLC circuit, an ac source on the left connected in a loop through a resistor R along the top, an inductor L on the right side, and a capacitor C along the bottom. Panel b: a phasor diagram for the case where the inductive reactance is larger than the capacitive reactance. The current phasor I sub m lies along the horizontal axis. The resistor's voltage phasor V sub R lies along it too. The inductor's voltage phasor V sub L points straight up and the capacitor's V sub C straight down, shorter. Their sum, V sub L minus V sub C, points up. The source voltage phasor V sub m is the diagonal of the rectangle formed by V sub R and V sub L minus V sub C, at angle phi above the current phasor.

Now connect a resistor, an inductor, and a capacitor in series with an ac source (Figure 26.5a). The same current $i = I_m \sin(\omega t - \phi)$ flows through all three elements; the source voltage is $v = V_m \sin \omega t$, and ϕ, the **phase constant**, is the angle by which the current lags the source voltage. We want I_m and ϕ.

By the loop rule, the source voltage equals the sum of the three voltages at every instant, $v = v_R + v_L + v_C$. But the three are not in phase, so their amplitudes do not simply add. This is where phasors pay off. Draw the current phasor I_m (Figure 26.5b). The resistor's voltage phasor, of length $V_R = I_m R$, is in phase with it. The inductor's, $V_L = I_m X_L$, is 90° ahead of it, and the capacitor's, $V_C = I_m X_C$, is 90° behind it. The instantaneous values add, so the phasors add as vectors, and the source phasor V_m is their vector sum. Since V_L and V_C point in opposite directions, they combine to $V_L - V_C$, at right angles to V_R, and by the Pythagorean theorem

$$V_m^2 = V_R^2 + (V_L - V_C)^2 = I_m^2 \left[R^2 + (X_L - X_C)^2\right]. \tag{26.5.1}$$

So $V_m = I_m Z$, where

Key Equation: Impedance and phase of a series RLC circuit

$$Z = \sqrt{R^2 + (X_L - X_C)^2}, \qquad \tan\phi = \frac{X_L - X_C}{R}. \tag{26.7}$$

The quantity Z is the **impedance** of the circuit, measured in ohms; it generalizes resistance to ac circuits, with $I_m = V_m/Z$ and $I_{\text{rms}} = V_{\text{rms}}/Z$. If $X_L > X_C$, the circuit is *inductive*: ϕ is positive, and the current lags the source voltage. If $X_C > X_L$, it is *capacitive*: ϕ is negative, and the current leads. A series circuit with only some of the elements is handled by setting the missing reactance or resistance to zero.

Example 26.6: A series RLC circuit

A series circuit has $R = 200\ \Omega$, $L = 0.40\ \text{H}$, and $C = 10\ \mu\text{F}$, driven by a $120\ \text{V}$ (rms), $60\ \text{Hz}$ source. Find (a) the reactances, (b) the impedance, (c) the rms current, (d) the phase constant, and (e) the rms voltage across each element.

Solution

(a) With $\omega = 2\pi(60) = 377\ \text{rad/s}$:

$$X_L = \omega L = (377)(0.40) = 151\,\Omega, \qquad X_C = \frac{1}{\omega C} = \frac{1}{(377)(10\times 10^{-6})} = 265\,\Omega. \tag{26.5.2}$$

(b) The impedance is

$$Z = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{(200)^2 + (151-265)^2} = \boxed{230\,\Omega}. \tag{26.5.3}$$

(c) $I_{\text{rms}} = V_{\text{rms}}/Z = 120/230 = \boxed{0.52\ \text{A}}$.

(d) $\tan\phi = (X_L - X_C)/R = -114/200 = -0.57$, so $\phi = \boxed{-30^\circ}$. The circuit is capacitive, and the current leads the source voltage by 30°.

(e) $V_R = I_{\text{rms}}R = \boxed{104\ \text{V}}$, $V_L = I_{\text{rms}}X_L = \boxed{79\ \text{V}}$, and $V_C = I_{\text{rms}}X_C = \boxed{138\ \text{V}}$. These add up to far more than $120\ \text{V}$, and V_C alone exceeds it. There is no contradiction: the voltages reach their peaks at different times, and it is their phasors that add, $\sqrt{104^2 + (79-138)^2} = 120\ \text{V}$.

Resonance

The current in a series RLC circuit, $I_{\text{rms}} = V_{\text{rms}}/Z$, depends on the driving frequency through the reactances. At low frequency the capacitor's reactance is huge; at high frequency the inductor's is. In between, at the frequency where they are equal, $X_L = X_C$, they cancel, the impedance takes its smallest possible value, $Z = R$, and the current is greatest. Setting $\omega L = 1/\omega C$ gives the **resonant frequency**:

Key Equation: Resonance in a series RLC circuit

$$\omega_0 = \frac{1}{\sqrt{LC}}, \qquad I_{\text{rms, max}} = \frac{V_{\text{rms}}}{R}. \tag{26.8}$$

This is exactly the natural frequency of the LC circuit of Section 26.1. When the source drives the circuit at its own natural frequency, the response is largest: this is **resonance**. At resonance, the phase constant is zero; the circuit behaves like the resistor alone, with the voltages across the inductor and capacitor, which can each be much larger than the source voltage, exactly canceling each other.

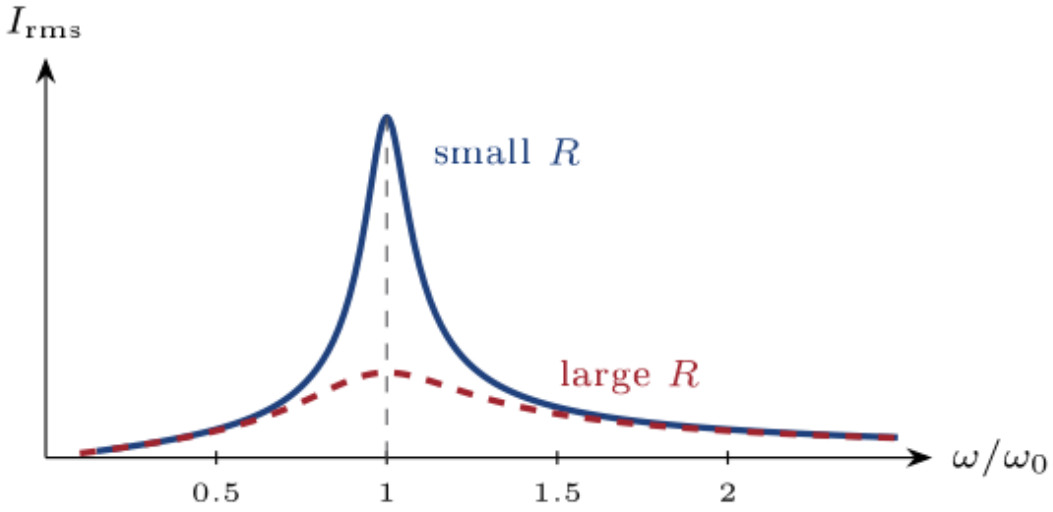


Figure 26.6. Resonance. The rms current in a series RLC circuit peaks at the resonant frequency $\omega_0 = 1/\sqrt{LC}$. With a smaller resistance (solid), the peak is higher and sharper.

•Figure description

A graph of the rms current in a series RLC circuit against the driving angular frequency omega, in units of the resonant frequency omega zero, from 0 to 2.5. Two curves are shown, for the same inductor and capacitor with two different resistances. Both peak at omega equals omega zero. The curve for the smaller resistance has a peak four times higher and much narrower; the curve for the larger resistance has a low, broad peak.

The smaller the resistance, the higher and sharper the peak in the current (Figure 26.6). A circuit with a sharp resonance responds strongly to frequencies close to ω_0 and hardly at all to others. That is how a radio selects a station: the signals of all the stations produce tiny emfs in the antenna, but only the one whose frequency matches the tuning circuit's resonant frequency drives a large current. The same physics, a driven oscillator responding most strongly at its natural frequency, is treated for mechanical systems in Section 28.9.

Example 26.7: Resonance

For the circuit of Example 26.6, find (a) the resonant frequency and (b) the rms current and the rms voltage across the capacitor at resonance.

Solution

(a) $\omega_0 = 1/\sqrt{LC} = 1/\sqrt{(0.40)(10 \times 10^{-6})} = 500 \text{ rad/s}$, so $f_0 = \omega_0/2\pi = \boxed{80 \text{ Hz}}$.

(b) At resonance $Z = R$, so $I_{\text{rms}} = 120/200 = \boxed{0.60 \text{ A}}$. The capacitor's reactance there is $X_C = 1/\omega_0 C = 200 \ \Omega$, so $V_C = (0.60)(200) = \boxed{120 \text{ V}}$. With a smaller resistance, say $20 \ \Omega$, the current at resonance would be 6.0 A and the capacitor's voltage 1200 V, ten times the source voltage. Resonant circuits can build up dangerously large voltages from modest sources, and their components must be rated for them.

Checkpoint 26.3

A series RLC circuit is driven below its resonant frequency. (a) Is its reactance mainly inductive or capacitive? (b) Does the current lead or lag the source voltage? (c) To bring it into resonance at this frequency, should you increase or decrease the capacitance?

Answer

Answers to Checkpoints

Checkpoint 26.3. (a) Below resonance, $X_C = 1/\omega C$ is larger than $X_L = \omega L$, so the reactance is mainly capacitive. (b) The current *leads* the source voltage, as it does in a purely capacitive circuit. (c) Increase it. Resonance occurs at $\omega_0 = 1/\sqrt{LC}$, so to lower ω_0 down to the driving frequency you must make C larger (which also makes X_C smaller, until it equals X_L).

This page titled 26.5: Series RLC Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.6: Power in AC Circuits

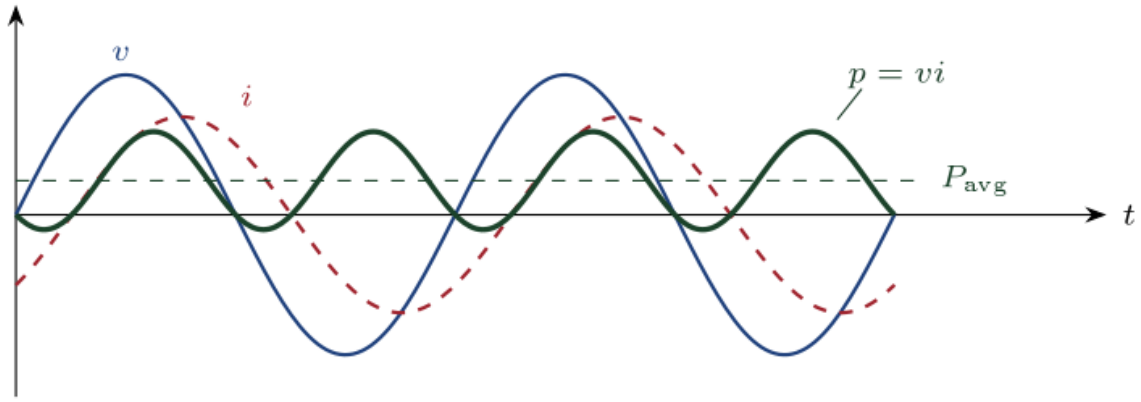


Figure 26.7. Instantaneous power in an ac circuit where the current lags the voltage by ϕ. The power $p = vi$ oscillates at twice the source frequency and is briefly negative, as the circuit's fields return energy to the source. Its average is $I_{\text{rms}}V_{\text{rms}}\cos\phi$.

•Figure description

Graphs against time of the voltage v, a sine curve; the current i, which lags the voltage by a phase angle phi; and the instantaneous power p, their product. The power curve oscillates at twice the frequency; it is mostly positive but dips below zero for part of each half cycle. A horizontal dashed line shows its average value, P average, which is positive but smaller than the peak.

The instantaneous power delivered by the source is $p = vi$. For a resistor, v and i are in phase, their product is never negative, and the average power is $I_{\text{rms}}^2 R$. For an ideal capacitor or inductor, v and i are 90° apart, and their product is positive half the time and negative half the time (Figure 26.7). Energy flows into the element's field for a quarter cycle and back out to the source the next, and the average power is zero. Capacitors and inductors store and return energy; only resistance dissipates it.

In a series RLC circuit, then, the average power is the power dissipated in the resistor, $P_{\text{avg}} = I_{\text{rms}}^2 R$. Since $R = Z\cos\phi$ (Figure 26.5b) and $I_{\text{rms}}Z = V_{\text{rms}}$, this can be written

Key Equation: Average power in an ac circuit

$$P_{\text{avg}} = I_{\text{rms}}^2 R = I_{\text{rms}}V_{\text{rms}}\cos\phi. \tag{26.9}$$

The factor $\cos\phi$ is the **power factor**. For a purely resistive circuit, or any circuit at resonance, $\phi = 0$ and the power factor is 1: the source delivers the full $I_{\text{rms}}V_{\text{rms}}$. For a circuit that is mainly reactive, $\cos\phi$ is small, and the source must supply a large current to deliver a modest power.

Example 26.8: Power in the RLC circuit

Find the average power delivered to the circuit of Example 26.6, and its power factor.

Solution

$P_{\text{avg}} = I_{\text{rms}}^2 R = (0.52)^2(200) = \boxed{54\ \text{W}}$. The power factor is $\cos\phi = R/Z = 200/230 = \boxed{0.87}$. As a check, $I_{\text{rms}}V_{\text{rms}}\cos\phi = (0.52)(120)(0.87) = 54\ \text{W}$. At resonance, with $I_{\text{rms}} = 0.60\ \text{A}$ and a power factor of 1, the power would be $(0.60)(120) = 72\ \text{W}$.

Example 26.9: Why utilities care about power factor

An industrial motor uses $10\ \text{kW}$ from a $240\ \text{V}$ (rms) supply, with a power factor of 0.70 (motors are inductive). (a) What current does it draw? (b) By what factor would the power lost in the supply lines drop if the power factor were raised to 1?

Solution

(a) From $P_{\text{avg}} = I_{\text{rms}} V_{\text{rms}} \cos\phi$,

$$I_{\text{rms}} = \frac{P_{\text{avg}}}{V_{\text{rms}} \cos\phi} = \frac{10\,000}{(240)(0.70)} = \boxed{60\ \text{A}}. \tag{26.6.1}$$

(b) With a power factor of 1, the current for the same power is $10\ 000/240 = 42\ \text{A}$. The line losses go as I^2, so they drop by the factor $(0.70)^{-2} = \boxed{2.0}$. The extra current at a low power factor does no useful work, but it heats the wires and ties up the capacity of generators and transformers. Utilities therefore charge industrial customers for low power factors, and factories install banks of capacitors beside their motors. The capacitors' leading current cancels the motors' lagging current, bringing the power factor close to 1.

Checkpoint 26.4

(a) A circuit contains only an ideal inductor and a capacitor. What average power does the source deliver? (b) An appliance is labeled $1200\ \text{W}$ and $12\ \text{A}$ at $120\ \text{V}$. What is its power factor?

Answer

Answers to Checkpoints

Checkpoint 26.4. (a) Zero. In ideal inductors and capacitors the current and voltage are 90° out of phase, so the power factor $\cos\phi$ is zero: energy flows into the elements for a quarter cycle and back to the source for the next, averaging to nothing. (b) The power factor is the real power divided by $V_{\text{rms}} I_{\text{rms}}$: $\cos\phi = 1200/[(120)(12)] = 1200/1440 \approx 0.83$.

This page titled 26.6: Power in AC Circuits is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.S: AC Circuits (Summary)

Summary

- In an **LC circuit**, charge and current oscillate with angular frequency $\omega = 1/\sqrt{LC}$, a quarter cycle apart, and the energy passes between the capacitor's electric field and the inductor's magnetic field. The equation is that of a mass on a spring.
- An ac source has $v = V_m \sin \omega t$. **Rms values**, $I_{\text{rms}} = I_m/\sqrt{2}$ and $V_{\text{rms}} = V_m/\sqrt{2}$, give the same average power in a resistor as a steady current. Sinusoids of the same frequency are represented by rotating **phasors**.
- In a resistor, current and voltage are in phase. In a capacitor, the current leads by 90°, with **capacitive reactance** $X_C = 1/\omega C$. In an inductor, the current lags by 90°, with **inductive reactance** $X_L = \omega L$.
- In a series RLC circuit, $I_m = V_m/Z$, with **impedance** $Z = \sqrt{R^2 + (X_L - X_C)^2}$, and the current lags the source by ϕ, with $\tan \phi = (X_L - X_C)/R$. The element voltages add as phasors.
- At **resonance**, $\omega_0 = 1/\sqrt{LC}$, $X_L = X_C$, $Z = R$, and the current is greatest; smaller R gives a sharper resonance.
- Only resistance dissipates power: $P_{\text{avg}} = I_{\text{rms}}^2 R = I_{\text{rms}} V_{\text{rms}} \cos \phi$, where $\cos \phi$ is the **power factor**.

Key Equations

LC oscillations

$$q = Q \cos(\omega t + \phi), \qquad \omega = \frac{1}{\sqrt{LC}} \tag{26.S.1}$$

Energy in an LC circuit

$$U = \frac{q^2}{2C} + \frac{1}{2} L i^2 = \frac{Q^2}{2C} = \frac{1}{2} L I^2 \tag{26.S.2}$$

Rms values

$$I_{\text{rms}} = \frac{I_m}{\sqrt{2}}, \qquad V_{\text{rms}} = \frac{V_m}{\sqrt{2}} \tag{26.S.3}$$

Reactances

$$X_C = \frac{1}{\omega C}, \qquad X_L = \omega L \tag{26.S.4}$$

Series RLC circuit

$$Z = \sqrt{R^2 + (X_L - X_C)^2}, \qquad \tan \phi = \frac{X_L - X_C}{R}, \qquad I_m = \frac{V_m}{Z} \tag{26.S.5}$$

Resonance

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{26.S.6}$$

Average power

$$P_{\text{avg}} = I_{\text{rms}}^2 R = I_{\text{rms}} V_{\text{rms}} \cos \phi \tag{26.S.7}$$

This page titled 26.S: AC Circuits (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

26.E: AC Circuits (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Voltages and currents are rms values unless stated otherwise.

Section 26.1 — LC Oscillations

26.1 ★ Find the frequency of oscillation of an LC circuit with $L = 50\ \text{mH}$ and $C = 2.0\ \mu\text{F}$.

26.2 ★★ A $4.0\ \mu\text{F}$ capacitor charged to $50\ \text{V}$ is connected across a $10\ \text{mH}$ inductor. Find (a) the maximum current, (b) the time for the capacitor to discharge completely for the first time, and (c) the charge on the capacitor when the current is half its maximum.

26.3 ★★ *[Conceptual.]* In an LC circuit, what is the frequency with which the energy stored in the capacitor oscillates? Why is it different from the frequency of the charge?

26.4 ★★★ Show that in an LC circuit the charge and current satisfy $q^2/Q^2 + i^2/I^2 = 1$ at all times, where Q and I are the amplitudes. Use it to find the charge at the moment when the energy is shared equally between the capacitor and the inductor.

Section 26.2 — Resistors in an AC Circuit

26.5 ★ The voltage of a European outlet is $230\ \text{V}$ rms. What is its amplitude?

26.6 ★ A $1500\ \text{W}$ heater runs on $120\ \text{V}$. Find its resistance and the rms and peak currents.

26.7 ★★ *[Conceptual.]* An ac ammeter in series with a resistor reads $2.0\ \text{A}$. A student says the current is always $2.0\ \text{A}$. Correct the statement, and give the largest value the current reaches.

Section 26.3 — Capacitors in an AC Circuit

26.8 ★ Find the reactance of a $0.47\ \mu\text{F}$ capacitor at $60\ \text{Hz}$ and at $1.0\ \text{MHz}$.

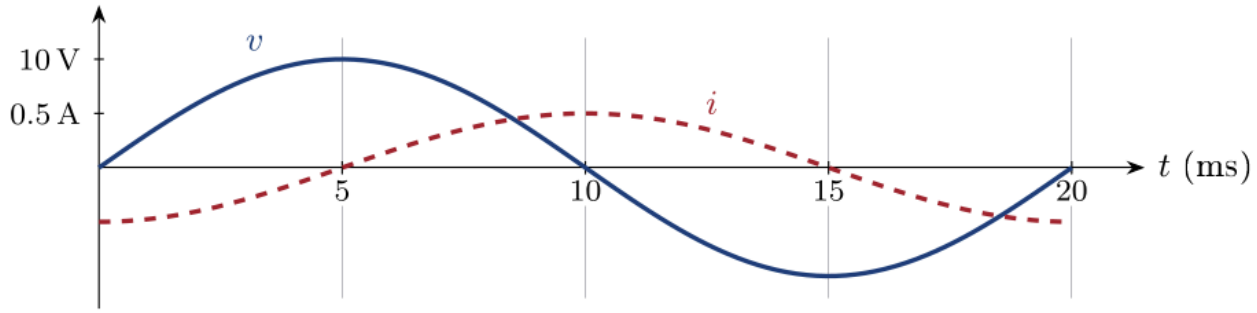
26.9 ★★ At what frequency does a $2.2\ \mu\text{F}$ capacitor have a reactance of $100\ \Omega$? What rms current does it pass when connected to $12\ \text{V}$ rms at that frequency?

Section 26.4 — Inductors in an AC Circuit

26.10 ★ Find the reactance of a $15\ \text{mH}$ inductor at $60\ \text{Hz}$ and at $20\ \text{kHz}$.

26.11 ★★ At what frequency do a $2.0\ \text{mH}$ inductor and a $5.0\ \mu\text{F}$ capacitor have the same reactance? What is that reactance?

26.12 ★★ *[Graph-reading.]* The graph shows the voltage across, and the current through, an unknown circuit element connected to an ac source.



•Figure description
A graph against time, in milliseconds from 0 to 20, of voltage in volts, a sine curve of amplitude 10 volts, and current in amperes, a curve of amplitude 0.5 amperes that reaches its peaks 5 milliseconds after the voltage does. One full cycle takes 20 milliseconds.

a. What is the frequency of the source?
b. Does the current lead or lag the voltage, and by what angle? Is the element a resistor, a capacitor, or an inductor?
c. Find its reactance and its inductance or capacitance.

Section 26.5 — Series RLC Circuits

26.13 ★ A series circuit has $R = 30\ \Omega$, $X_L = 80\ \Omega$, and $X_C = 40\ \Omega$. Find the impedance and the phase constant.

26.14 ★★ A $60\ \Omega$ resistor and a $0.20\ \text{H}$ inductor are in series with a $120\ \text{V}$, $60\ \text{Hz}$ source. Find the current, the phase constant, and the voltages across the resistor and the inductor. Check that the voltages add as phasors to $120\ \text{V}$.

26.15 ★★ *[Conceptual.]* In a series RLC circuit, the rms voltage across the capacitor is larger than the rms voltage of the source. How is this possible without violating energy conservation?

26.16 ★★★ A coil of unknown resistance and inductance draws $2.0\ \text{A}$ from a $12\ \text{V}$ battery, and $1.2\ \text{A}$ from a $12\ \text{V}$ rms, $60\ \text{Hz}$ source. Find its resistance and inductance.

Section 26.5.1 — Resonance

26.17 ★ What capacitance must be used with a $3.0\ \mu\text{H}$ inductor to make a circuit resonate at $100\ \text{MHz}$, in the FM radio band?

26.18 ★★ A series RLC circuit has $R = 10\ \Omega$, $L = 20\ \text{mH}$, and $C = 0.50\ \mu\text{F}$, with a $5.0\ \text{V}$ source. Find (a) the resonant frequency, (b) the current at resonance, and (c) the rms voltage across the inductor at resonance.

26.19 ★★ *[Real-world: radio.]* Two radio stations broadcast at $980\ \text{kHz}$ and $1000\ \text{kHz}$. Explain why a tuning circuit needs a small resistance to separate them, using Figure 26.6.

26.20 ★★★ Show that at frequencies well above resonance, the current in a series RLC circuit is approximately $V/\omega L$, and well below it, approximately $V\omega C$. Sketch $I(\omega)$ on logarithmic axes.

Section 26.6 — Power in AC Circuits

26.21 ★ A circuit draws $5.0\ \text{A}$ from a $120\ \text{V}$ source with a power factor of 0.80. Find the average power.

26.22 ★★ For the circuit of Problem 26.14, find the average power and the power factor. How much capacitance, placed in series, would raise the power factor to 1 at $60\ \text{Hz}$?

26.23 ★★ *[Real-world: energy.]* An electricity meter records average power. A customer argues that a capacitor bank that "stores and returns" energy should not change her bill. Is she right? Whose bill does the capacitor bank change, and why?

26.24 ★★★ Show that the average power delivered to a series RLC circuit, as a function of driving frequency, is P

$= V_{\text{rms}}^2 R/Z^2$, and that it falls to half its maximum value when $X_L - X_C = \pm R$.

This page titled 26.E: AC Circuits (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27: ELECTROMAGNETIC WAVES

In 2019 a spacecraft the size of a loaf of bread unfurled a sheet of shimmering Mylar the area of a boxing ring and began, very gently, to accelerate. It carried no engine, burned no fuel, and ejected nothing at all. Its only push came from sunlight bouncing off the sail. This was *LightSail 2*, and over the following months it raised its orbit measurably on sunlight alone. A beam of light, it turns out, can shove a spacecraft across the solar system.

How can light push on anything? The answer is one of the crowning achievements of classical electromagnetism. Over the past several chapters you have met four laws: Gauss's law for electric fields, its magnetic counterpart, Faraday's law of induction, and Ampère's law. In the 1860s James Clerk Maxwell noticed that one of them was incomplete. When he repaired it, the four laws together predicted something no one had suspected: a self-sustaining wave of electric and magnetic fields, racing through empty space at a speed he could calculate from tabletop measurements. The speed came out equal to the measured speed of light. Light, Maxwell concluded, is an electromagnetic wave.

This chapter follows that argument from the missing term to its consequences. Electromagnetic waves carry energy, which is how the Sun warms the Earth. They carry momentum, which is what pushes a solar sail. And they come in an enormous range of wavelengths, from radio to gamma rays, of which visible light is only a sliver. By the end you will be able to compute the push of sunlight on a sail yourself.

Learning Objectives

- Explain why Ampère's law needs a displacement-current term, and use it to find the magnetic field inside a charging capacitor.
- State Maxwell's four equations and describe what each one says physically.
- Show that Maxwell's equations predict electromagnetic waves, which travel at the speed $c = 1/\sqrt{\mu_0 \epsilon_0}$, and relate the electric and magnetic fields of such a wave.
- Use the Poynting vector and intensity to describe the energy a wave transports, relating S, I, E, and B.
- Compute the radiation pressure light exerts on absorbing and reflecting surfaces.
- Locate the regions of the electromagnetic spectrum and convert between wavelength and frequency.

We will return to the solar sail, and finish the calculation of how hard sunlight can push, in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

27.1: Displacement Current
27.2: Maxwell's Equations
27.3: Propagation of Electromagnetic Waves
27.4: Energy Carried by Electromagnetic Waves
27.5: Momentum and Radiation Pressure of Electromagnetic Waves
27.6: The Electromagnetic Spectrum
27.S: Electromagnetic Waves (Summary)
27.E: Electromagnetic Waves (Problems)

This page titled 27: Electromagnetic Waves is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.1: Displacement Current

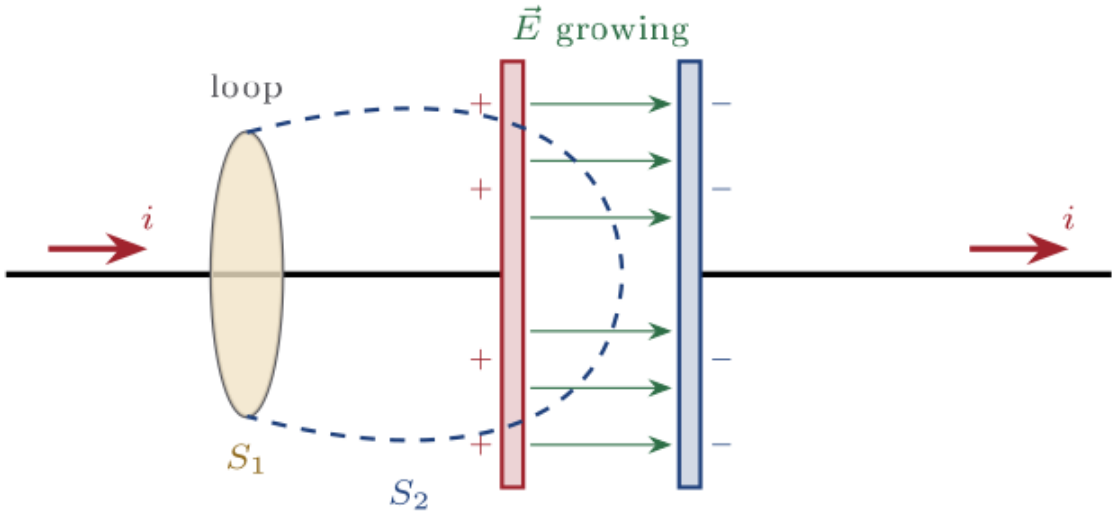


Figure 27.1. Ampère's original law fails for a charging capacitor. The loop around the wire bounds the flat surface S_1, which the current pierces, and the bulging surface S_2, which passes between the plates where no charge flows. Through S_2 passes only a changing electric field.

•Figure description

A parallel-plate capacitor being charged, drawn side-on with two vertical plates. A wire from the left carries current i into the left plate, and a wire on the right carries current i away from the right plate. Between the plates, horizontal electric field lines point from the left plate to the right plate and are growing stronger. A loop encircles the wire to the left of the capacitor. Surface S sub 1 is a flat disk spanning the loop, pierced by the wire. Surface S sub 2 has the same boundary loop but bulges to the right into the gap between the plates, so no wire passes through it; only the changing electric field does.

Ampère's law (Chapter 23) relates the magnetic field around a closed loop to the current passing through it:

$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 i_{\text{enc}}. \tag{27.1.1}$$

Here i_{enc} means the current through *any* surface bounded by the loop. For a steady current in a wire, every such surface is pierced by the same wire, and the choice does not matter. But consider a capacitor that is charging (Figure 27.1). A current i flows in the wire toward one plate and away from the other, yet no charge crosses the gap between the plates. Draw a loop around the wire. For the flat surface S_1 bounded by that loop, the wire passes through and $i_{\text{enc}} = i$. For the bulging surface S_2, which has the same boundary but passes between the plates, nothing passes through, and $i_{\text{enc}} = 0$. Ampère's law gives two different answers for the same loop. It cannot be complete.

What *does* pass through S_2 is electric field, and it is changing. As charge q builds up on plates of area A, the field between them grows as $E = q/\epsilon_0 A$ (Section 16.3), so the electric flux through the gap is $\Phi_E = EA = q/\epsilon_0$. Its rate of change is

$$\frac{d\Phi_E}{dt} = \frac{1}{\epsilon_0}\frac{dq}{dt} = \frac{i}{\epsilon_0}. \tag{27.1.2}$$

The quantity $\epsilon_0\ d\Phi_E/dt$ is therefore exactly equal to the current i in the wire. Maxwell called it the **displacement current**,

Key Equation: Displacement current

$$i_d = \epsilon_0 \frac{d\Phi_E}{dt}, \tag{27.1}$$

and added it to the current in Ampère's law. With it, the surface S_1 encloses a conduction current i and the surface S_2 encloses a displacement current $i_d = i$, and the two agree. The repaired law is the **Ampère–Maxwell law**:

$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0\left(i_{\text{enc}} + \epsilon_0\frac{d\Phi_E}{dt}\right). \tag{27.2}$$

Despite its name, the displacement current is not a flow of charge. It is a changing electric field, and the new law says that a changing electric field produces a magnetic field, exactly as a current does. This is the mirror image of Faraday's law (Chapter 24), in which a changing magnetic field produces an electric field. That symmetry is about to have dramatic consequences.

Example 27.1: The magnetic field inside a charging capacitor

A capacitor with circular plates of radius $R = 5.0\ \text{cm}$ is being charged by a steady current of $2.0\ \text{A}$. (a) At what rate is the electric field between the plates changing? (b) Find the magnetic field between the plates at a distance $r = 3.0\ \text{cm}$ from the central axis, and (c) outside the plates at $r = 10\ \text{cm}$.

Solution

(a) The field between the plates is $E = q/\epsilon_0 A$, so

$$\frac{dE}{dt} = \frac{i}{\epsilon_0 A} = \frac{2.0}{(8.85\times10^{-12})\pi(0.050)^2} = \boxed{2.9\times10^{13}\,\text{V/(m}\cdot\text{s)}}. \tag{27.1.3}$$

(b) By symmetry, the magnetic field circles the axis and has the same magnitude all around a circle of radius r. Apply the Ampère–Maxwell law to that circle. There is no conduction current between the plates, but the displacement current is spread uniformly over the plate area, so a circle of radius $r < R$ encloses the fraction r^2/R^2 of it:

$$B(2\pi r) = \mu_0\, i\,\frac{r^2}{R^2} \qquad\Longrightarrow\qquad B = \frac{\mu_0 i\, r}{2\pi R^2}. \tag{27.1.4}$$

Numerically,

$$B = \frac{(1.257\times10^{-6})(2.0)(0.030)}{2\pi(0.050)^2} = \boxed{4.8\times10^{-6}\,\text{T}}. \tag{27.1.5}$$

(c) Outside the plates, $r > R$, the circle encloses the whole displacement current i, and

$$B = \frac{\mu_0 i}{2\pi r} = \frac{(1.257\times10^{-6})(2.0)}{2\pi(0.10)} = \boxed{4.0\times10^{-6}\,\text{T}}, \tag{27.1.6}$$

exactly the field the same current would produce around a long straight wire. From outside, the gap between the plates is magnetically invisible: the displacement current continues the conduction current seamlessly across it.

Checkpoint 27.1

A capacitor is charging. An Ampèrian loop encircles the wire leading to one plate; an identical loop, with the same boundary, instead passes between the plates where no charge flows. Ampère's *original* law gives different magnetic fields for the two loops. Which term in the Ampère–Maxwell law repairs the contradiction, and what physical quantity plays the role of the "missing" current between the plates? What happens to the displacement current once the capacitor is fully charged?

Answer

Answers to Checkpoints

Checkpoint 27.1. The displacement-current term, $\mu_0\epsilon_0\ d\Phi_E/dt$, repairs it. Between the plates no charge flows, but the electric field there is growing, and the displacement current $i_d = \epsilon_0\ d\Phi_E/dt$ turns out to equal the conduction current in the wire exactly. Both loops then enclose the same total current and give the same magnetic field. Once the capacitor is fully charged, the field between the plates stops changing, so $i_d = 0$, and the conduction current in the wire has also dropped to zero, so the two loops still agree.

This page titled 27.1: Displacement Current is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.2: Maxwell's Equations

With the displacement current in place, the laws of classical electricity and magnetism are complete. Four equations, together called **Maxwell's equations**, describe every electric and magnetic field, whether static or changing:

Key Equation: Maxwell's equations (integral form)

$$\oint \mathbf{E}\cdot d\mathbf{A} = \frac{q_{\text{enc}}}{\epsilon_0} \qquad \text{(Gauss, electric)} \tag{27.3}$$

$$[2pt] \oint \mathbf{B}\cdot d\mathbf{A} = 0 \qquad \text{(Gauss, magnetic)} \tag{27.4}$$

$$[2pt] \oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt} \qquad \text{(Faraday)} \tag{27.5}$$

$$[2pt] \oint \mathbf{B}\cdot d\mathbf{s} = \mu_0 i_{\text{enc}} + \mu_0\epsilon_0\frac{d\Phi_E}{dt} \qquad \text{(Ampère--Maxwell)} \tag{27.6}$$

Each says something simple.

- **Gauss's law for electric fields** (Chapter 17): electric field lines begin and end on charges, so the electric flux out of a closed surface measures the charge inside.
- **Gauss's law for magnetic fields** (Chapter 23): the magnetic flux out of any closed surface is zero. Magnetic field lines never begin or end; they form closed loops. This is the mathematical statement that there are no magnetic monopoles, no isolated north or south poles to act as sources.
- **Faraday's law of induction** (Chapter 24): a changing magnetic flux produces a circulating electric field.
- **The Ampère–Maxwell law**: a current, or a changing electric flux, produces a circulating magnetic field.

Together with the Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ (Chapter 22), which says what the fields do to charges, they contain all of classical electromagnetism, from the spark of static electricity to the operation of a transformer.

Note: Maxwell's equations in differential form

The integral forms describe fields over whole surfaces and loops. Shrink the surface or the loop down around a single point, and the same four laws become statements about the fields *at that point*. This *differential form* uses two operations from vector calculus. The *divergence* $\nabla \cdot \mathbf{E}$ measures how strongly a field spreads outward from a point, like water from a spring: it is the flux out of a tiny closed surface around the point, divided by the volume inside. The *curl* $\nabla \times \mathbf{E}$ measures how strongly a field circulates around a point, like water in an eddy: its component along any direction is the circulation around a tiny loop perpendicular to that direction, divided by the loop's area. With ρ the charge per unit volume and $\mathbf{J}$ the current density (Chapter 20), Maxwell's equations read

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \text{(Gauss, electric)} \tag{27.7}$$

$$[2pt]\nabla \cdot \mathbf{B} = 0 \qquad \text{(Gauss, magnetic)} \tag{27.8}$$

$$[2pt]\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(Faraday)} \tag{27.9}$$

$$[2pt]\nabla \times \mathbf{B} = \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampère–Maxwell)} \tag{27.10}$$

Each line says the same thing as its integral partner: charge is where electric field lines diverge from; magnetic field lines diverge from nowhere; a changing magnetic field makes the electric field curl; and a current or a changing electric field makes the magnetic field curl. The last term of the last line is the one Maxwell added, and it is from this term that light is born. You will not need the differential form in this course, but upper-division courses in electromagnetism use it almost exclusively.

Note: Why the missing term was never noticed

If Maxwell's term is so important, why did Ampère and Faraday, careful experimenters both, never see its effects? Because in the circuits they could build it is tiny. The factor $\mu_0\epsilon_0 = 1/c^2 \approx 1.1 \times 10^{-17}\ \mathrm{s^2/m^2}$ is extraordinarily small, so a changing electric field produces a noticeable magnetic field only when it changes extremely fast. In Example 27.1 the field between the plates had to change at some 10^{13} volts per meter every second to match an ordinary two-ampere current. Maxwell found the term not by measurement but by insisting that the equations be consistent, and its consequences were confirmed only in 1887, when Heinrich Hertz generated and detected the first radio waves.

Checkpoint 27.2

Look at the two Gauss's laws and the two circulation laws (Faraday and Ampère–Maxwell). (a) Which of the four would change if a magnetic monopole were discovered, and how? (b) In empty space, where $q_{\mathrm{enc}} = 0$ and $i_{\mathrm{enc}} = 0$, write the two circulation laws and describe the symmetry between them. What is still different about them?

Answer

Answers to Checkpoints

Checkpoint 27.2. (a) Gauss's law for magnetism would change: $\oint \mathbf{B} \cdot d\mathbf{A}$ would equal μ_0 times the enclosed magnetic charge instead of zero. To keep the equations consistent, Faraday's law would also gain a "magnetic current" term, mirroring the conduction current in the Ampère–Maxwell law. (b) In empty space, $\oint \mathbf{E} \cdot d\mathbf{s} = -\ d\Phi_B/dt$ and $\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0\epsilon_0\ d\Phi_E/dt$. Each says that a changing flux of one field creates a circulating version of the other: the mutual generation that makes a self-sustaining electromagnetic wave possible. They still differ in two ways: the minus sign in Faraday's law (Lenz's law) is absent from the other, and the constant $\mu_0\epsilon_0 = 1/c^2$ appears in only one of them.

This page titled 27.2: Maxwell's Equations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.3: Propagation of Electromagnetic Waves

In empty space, far from any charges or currents, Maxwell's equations still allow a changing magnetic field to create an electric field (Faraday) and a changing electric field to create a magnetic field (Ampère–Maxwell). Each field sustains the other. Once started, such a disturbance can travel on its own, needing no charges and no medium at all. To see what it looks like, we need a few facts about waves.

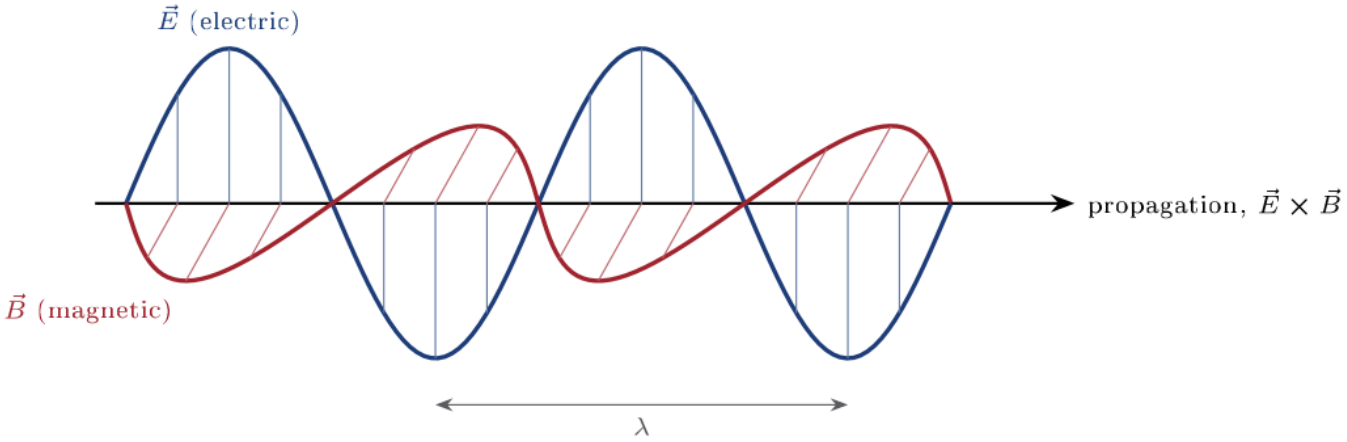


Figure 27.2. An electromagnetic wave. The electric field $\mathbf{E}$ (blue) and magnetic field $\mathbf{B}$ (red) oscillate in phase, perpendicular to each other and to the direction of travel. The wave advances along $\mathbf{E} \times \mathbf{B}$. One wavelength λ is marked.

•Figure description

A three-dimensional sketch of an electromagnetic wave traveling to the right along a horizontal axis. A blue sine curve labeled E, the electric field, oscillates up and down in the vertical plane, with short stems joining each point of the curve to the axis. A red sine curve labeled B, the magnetic field, oscillates in the horizontal plane at right angles to it, drawn in perspective so that it swings toward and away from the viewer; its label sits at the lower left, beside the first point where it swings toward the viewer. The two curves peak and cross zero at the same places, so the fields are in phase, and each is perpendicular to the other and to the direction of travel, which is labeled E cross B. A double-headed arrow beneath, from one trough of the electric field to the next, marks one wavelength lambda.

Traveling waves. A **wave** is a disturbance that moves from place to place, carrying energy without carrying matter along with it. Chapter 29 studies waves on strings and Chapter 30 sound waves in detail. Here we need only the simplest kind, a sinusoidal wave traveling in the $+x$ direction:

$$y(x,t) = y_m \sin(kx - \omega t). \tag{27.11}$$

At any instant the disturbance repeats every **wavelength** λ along x, and at any point it repeats every period $T = 1/f$ in time. The **angular wave number** $k = 2\pi/\lambda$ and the angular frequency $\omega = 2\pi f$ package those facts. Because the pattern depends on x and t only through the combination $kx - \omega t$, a crest keeps a constant value of $kx - \omega t$ as time goes on, and therefore moves forward at speed

$$v = \frac{\omega}{k} = f\lambda. \tag{27.3.1}$$

Every such wave also satisfies the **wave equation**,

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\,\frac{\partial^2 y}{\partial t^2}, \tag{27.12}$$

as you can check by differentiating equation (27.11) twice with respect to x and twice with respect to t. The wave equation is the signature of a wave: whenever the laws of physics lead to an equation of this form, a wave is the result. It also tells us the wave's speed. In equation (27.12) the number multiplying the time derivative is $1/v^2$, so if some law produces this equation with a different constant in that spot, setting the constant equal to $1/v^2$ gives the speed of the wave.

Maxwell's wave. Now apply Faraday's law and the Ampère–Maxwell law to fields in empty space. Restrict attention to a wave traveling along x with $\mathbf{E}$ along y and $\mathbf{B}$ along z, the arrangement drawn in Figure 27.2. Applying the two circulation laws to small rectangular loops, one in the xy-plane and one in the xz-plane, reduces them to

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t}, \qquad -\frac{\partial B}{\partial x} = \mu_0\epsilon_0 \frac{\partial E}{\partial t}. \tag{27.13}$$

Differentiating the first with respect to x and the second with respect to t, and eliminating B, gives

$$\frac{\partial^2 E}{\partial x^2} = \mu_0\epsilon_0 \frac{\partial^2 E}{\partial t^2}. \tag{27.14}$$

This is the wave equation (27.12), and comparing the two shows that the wave speed is $v = 1/\sqrt{\mu_0\epsilon_0}$. Plug in the constants of electricity and magnetism and out drops a number Maxwell recognized at once:

Key Equation: Speed of light in vacuum

$$c = \frac{1}{\sqrt{\mu_0\epsilon_0}} = 2.998 \times 10^8 \text{ m/s} \approx 3.00 \times 10^8 \text{ m/s}. \tag{27.15}$$

This was the thunderclap. Two constants measured in tabletop experiments with capacitors and coils, the permittivity $\epsilon_0 = 8.85 \times 10^{-12}\ \text{C}^2/(\text{N}\cdot\text{m}^2)$ and the **permeability of free space** $\mu_0 = 1.257 \times 10^{-6}\ \text{T}\cdot\text{m/A}$, combine to give precisely the measured speed of light. Light *is* an electromagnetic wave. Radio waves, microwaves, and x-rays are electromagnetic waves too, all traveling at the same speed c. What sets visible light apart is only its wavelength, which happens to fall in the narrow range our eyes can detect (Section 27.6).

A traveling **electromagnetic wave** has the form

$$E = E_m \sin(kx - \omega t), \qquad B = B_m \sin(kx - \omega t). \tag{27.16}$$

The electric and magnetic fields are in phase, mutually perpendicular, and both perpendicular to the direction of travel (Figure 27.2): the wave is **transverse**. The direction of propagation is given by the cross product $\mathbf{E} \times \mathbf{B}$. Substituting the sinusoids into the first of equations (27.13) gives $kE_m = \omega B_m$, so the magnitudes of the fields are locked together by the speed of light:

Key Equation: Fields of an electromagnetic wave

$$\frac{E}{B} = c. \tag{27.17}$$

The electric field of a light wave, in V/m, is therefore numerically about 3×10^8 times larger than its magnetic field in teslas, which is why we usually speak of light's electric field.

Example 27.2: The fields of a radio wave

Far from a transmitter, a radio wave traveling in the $+x$ direction has an electric field of amplitude $E_m = 0.050\ \text{V/m}$ pointing along $\pm y$. (a) What is the amplitude of its magnetic field? (b) Along which axis does the magnetic field point?

Solution

(a) From equation (27.17),

$$B_m = \frac{E_m}{c} = \frac{0.050}{3.00 \times 10^8} = \boxed{1.7 \times 10^{-10}\ \text{T}}. \tag{27.3.2}$$

(b) The magnetic field is perpendicular to both $\mathbf{E}$ and the direction of travel, so it lies along $\pm z$. Checking the direction: when $\mathbf{E}$ points along $+y$, the cross product $\mathbf{E} \times \mathbf{B}$ must point along $+x$, and $\hat{\jmath} \times \hat{k} = \hat{\imath}$, so at that instant $\mathbf{B}$ points along $\boxed{+z}$. Half a cycle later both fields have reversed, and $\mathbf{E} \times \mathbf{B}$ still points along $+x$.

Checkpoint 27.3

At some instant, the electric field of a light wave at a certain point points north and its magnetic field points up. (a) In which direction is the wave traveling? (b) If the electric field's magnitude at that point is $600\ \mathrm{V/m}$, what is the magnetic field's magnitude?

Answer

Answers to Checkpoints

Checkpoint 27.3. (a) A wave travels in the direction of $\mathbf{E} \times \mathbf{B}$. Take east, north, and up as $+x$, $+y$, and $+z$: then $\hat{j} \times \hat{k} = \hat{\imath}$, so the wave travels due *east*. (b) $B = E/c = (600\ \mathrm{V/m})/(3.00 \times 10^8\ \mathrm{m/s}) = 2.0 \times 10^{-6}\ \mathrm{T}$, a field far weaker than the Earth's.

This page titled 27.3: Propagation of Electromagnetic Waves is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.4: Energy Carried by Electromagnetic Waves

Every wave does one essential job: it transports energy from one place to another without transporting matter. An electromagnetic wave carries its energy in its electric and magnetic fields, and the rate of that transport is described by the **Poynting vector** $\mathbf{S}$, defined as

$$\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E} \times \mathbf{B}. \tag{27.18}$$

Its direction is the direction of energy flow (the same $\mathbf{E} \times \mathbf{B}$ along which the wave travels), and its magnitude is the power per unit area, in watts per square meter. Because $\mathbf{E}$ and $\mathbf{B}$ are perpendicular, the magnitude is simply $S = EB/\mu_0$, and using $E = cB$ from equation (27.17) we can write it with either field alone:

$$S = \frac{1}{\mu_0}EB = \frac{1}{c\mu_0}E^2 = \frac{c}{\mu_0}B^2. \tag{27.19}$$

The Poynting vector is an *instantaneous* quantity, and since E^2 oscillates between zero and its peak, S flickers far too fast to measure directly. What instruments (and eyes) respond to is its time average, the **intensity** I of the wave: the average power delivered per unit area of a surface facing the wave. Averaging $\sin^2$ over a cycle gives a factor of $\frac{1}{2}$, conveniently absorbed by writing the fields as root-mean-square values, $E_{\text{rms}} = E_m/\sqrt{2}$, just as for the alternating currents of Chapter 26:

Key Equation: Intensity of an electromagnetic wave

$$I = S_{\text{avg}} = \frac{1}{c\mu_0}E_{\text{rms}}^2 = \frac{c}{\mu_0}B_{\text{rms}}^2 = \frac{E_m^2}{2c\mu_0}. \tag{27.20}$$

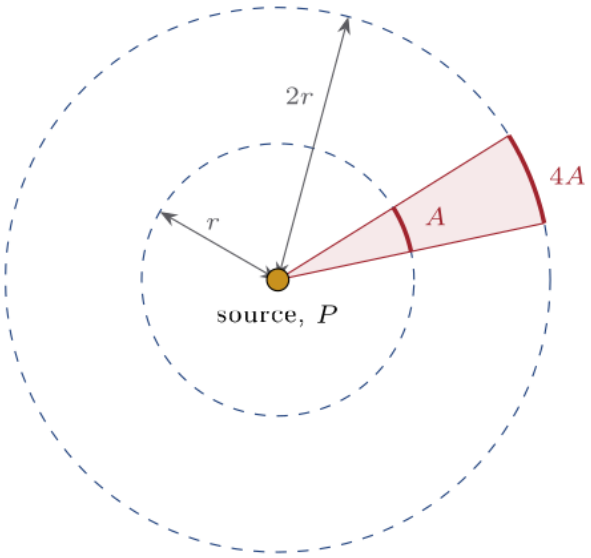


Figure 27.3. The inverse-square law for a point source. The power passing through area A at distance r spreads over area $4A$ at distance $2r$, so the intensity falls to a quarter.

•Figure description

A small point source at the center, labeled source, P, radiates equally in all directions. Two concentric dashed spheres surround it, with gray arrows from the source marking their radii, r for the inner sphere and 2 r for the outer. A shaded wedge bounded by two straight rays from the source cuts a patch of area A on the inner sphere; the same rays spread out to cut a patch four times larger, of area 4 A, on the outer sphere. The same power therefore passes through four times the area, so the intensity is one quarter as large.

For a source that radiates power P uniformly in all directions, that power spreads over the surface of an ever-growing sphere of area $4\pi r^2$, so the intensity falls off as the inverse square of the distance, for the same geometric reason as the gravitational field

of Chapter 13:

$$I = \frac{P}{4\pi r^2}. \tag{27.21}$$

Figure 27.3 shows the geometry. A beam that does not spread, such as a laser's, avoids this dilution almost entirely, which is why a feeble laser can be far more intense than a powerful lamp.

Example 27.3: The fields in sunlight

At the top of Earth's atmosphere, sunlight arrives with an intensity of about $I = 1.36 \times 10^3\ \mathrm{W/m^2}$ (the "solar constant"). Find the rms electric field and rms magnetic field of this light.

Solution

Solve equation (27.20) for the electric field. The combination $c\mu_0 = (3.00 \times 10^8)(1.257 \times 10^{-6}) = 377\ \Omega$ is called the impedance of free space:

$$E_{\mathrm{rms}} = \sqrt{I\, c\mu_0} = \sqrt{(1.36 \times 10^3)(377)} = \boxed{716\,\mathrm{V/m}}. \tag{27.4.1}$$

The magnetic field follows from $E = cB$:

$$B_{\mathrm{rms}} = \frac{E_{\mathrm{rms}}}{c} = \frac{716}{3.00 \times 10^8} = \boxed{2.39 \times 10^{-6}\,\mathrm{T}}. \tag{27.4.2}$$

The magnetic field is only a couple of microtesla, weaker than a refrigerator magnet, yet it is an equal partner with the electric field in carrying the Sun's energy. The factor of c between them is exactly what hides $\mathbf{B}$ from everyday notice.

Example 27.4: A laser pointer versus a light bulb

Compare the intensity of (a) a $1.0\ \mathrm{mW}$ laser pointer whose beam has a cross-sectional area of $1.0\ \mathrm{mm^2}$ with (b) the light from a bulb that radiates $100\ \mathrm{W}$ of electromagnetic power uniformly in all directions, at a distance of $1.0\ \mathrm{m}$. (c) Find the amplitude of the electric field in the laser beam.

Solution

(a) The laser's power is confined to its beam, so

$$I = \frac{P}{A} = \frac{1.0 \times 10^{-3}\,\mathrm{W}}{1.0 \times 10^{-6}\,\mathrm{m^2}} = \boxed{1.0 \times 10^3\,\mathrm{W/m^2}}. \tag{27.4.3}$$

(b) The bulb's power spreads over a sphere, so by equation (27.21)

$$I = \frac{P}{4\pi r^2} = \frac{100}{4\pi(1.0)^2} = \boxed{8.0\,\mathrm{W/m^2}}. \tag{27.4.4}$$

The one-milliwatt laser is more than a hundred times as intense as the hundred-watt bulb, and about as intense as full sunlight. That is why a laser pointer must never be aimed at anyone's eye.

(c) From equation (27.20), $E_m = \sqrt{2c\mu_0 I} = \sqrt{2(377)(1.0 \times 10^3)} = \boxed{870\,\mathrm{V/m}}$.

Checkpoint 27.4

A streetlight looks four times brighter to you than to a friend standing farther down the road. Ignoring obstructions, how many times farther from the lamp is your friend? (Use the inverse-square law.)

Answer

Answers to Checkpoints

Checkpoint 27.4. Intensity from a point source falls as $1/r^2$. The lamp looks four times as bright to you as to your friend, so your friend's distance squared is four times yours: your friend is $\sqrt{4} = 2$ times farther away.

This page titled 27.4: Energy Carried by Electromagnetic Waves is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.5: Momentum and Radiation Pressure of Electromagnetic Waves

An electromagnetic wave carries not only energy but *momentum*. This is surprising, since you learned momentum as $\mathbf{p} = m\mathbf{v}$ and a light wave has no mass, but Maxwell's equations require it. Picture the wave striking a surface that absorbs it. The wave's electric field drives the charges in the surface back and forth along $\mathbf{E}$. As they move, the wave's magnetic field exerts a force $q\mathbf{v} \times \mathbf{B}$ on them (Chapter 22), and a little thought shows that this force always points along $\mathbf{E} \times \mathbf{B}$, the direction the wave is traveling, whichever way the charges happen to be moving at the moment. The wave pushes the surface forward. Working out the size of the push shows that when a wave delivers energy U to an absorbing surface, it also delivers momentum

$$p = \frac{U}{c}. \tag{27.22}$$

Chapter 35 reaches the same result from a completely different direction: relativity's energy–momentum relation, $E^2 = (pc)^2 + (mc^2)^2$, reduces to $p = E/c$ for a massless photon. A rate of momentum delivery is a force, and spread over the illuminated area, that force is a pressure, the **radiation pressure**.

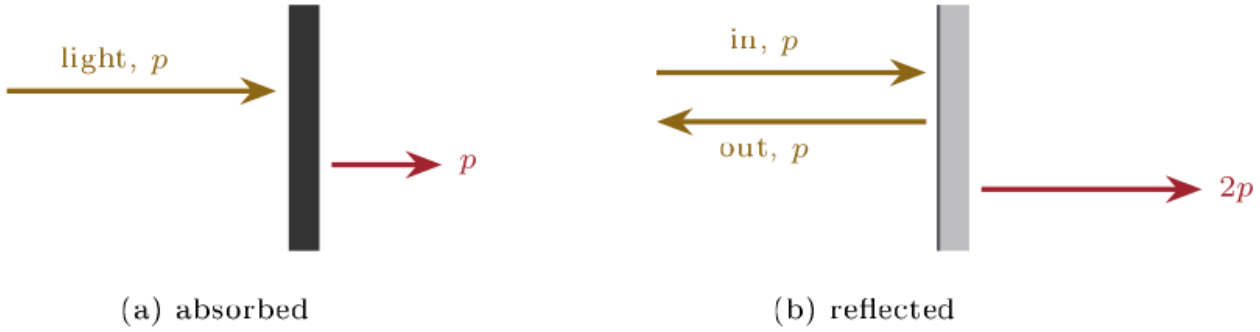


Figure 27.4. Momentum delivered by light. (a) An absorbing surface stops the light and receives its momentum p. (b) A mirror reverses the light's momentum, a change of $2p$, and so receives twice as much. The radiation pressure on a mirror is therefore twice that on an absorber.

•Figure description

Two panels comparing the momentum delivered by light. In panel a, light carrying momentum p strikes a black absorbing surface and stops; the surface receives momentum p, drawn as one arrow. In panel b, the same light strikes a mirror and bounces straight back carrying momentum p in the opposite direction; the change in the light's momentum is 2 p, and the mirror receives momentum 2 p, drawn as an arrow twice as long.

How much pressure depends on what the surface does with the light. If the surface *absorbs* the light completely, it receives all the incoming momentum. If instead it *reflects* the light straight back, the light's momentum reverses, so the surface receives twice as much (Figure 27.4), for the same reason that a ball bouncing off a wall pushes on it harder than a ball that sticks (Chapter 9). Since intensity is energy per unit time per unit area, dividing by c gives momentum per unit time per unit area, which is pressure:

Key Equation: Radiation pressure

$$P_r = \frac{I}{c} \text{ (total absorption)}, \qquad P_r = \frac{2I}{c} \text{ (total reflection)}. \tag{27.23}$$

The force on an illuminated area A is then $F = P_r A$. The numbers are tiny by everyday standards, which is why you do not feel sunlight pushing on your hand. But in the frictionless vacuum of space, applied continuously over months, that feeble push is enough to sail.

Example 27.5: The push on a small mirror

A perfectly reflecting square mirror $1.0\ \text{m}$ on a side is held face-on to sunlight at Earth's distance, where $I = 1.36 \times 10^3\ \text{W/m}^2$. Find the radiation pressure and the total force on the mirror.

Solution

The mirror reflects, so we use the reflection form of equation (27.23):

$$P_r = \frac{2I}{c} = \frac{2(1.36 \times 10^3)}{3.00 \times 10^8} = \boxed{9.1 \times 10^{-6}\,\text{Pa}}. \tag{27.5.1}$$

With area $A = (1.0)^2 = 1.0\ \text{m}^2$, the force is

$$F = P_r A = (9.1 \times 10^{-6})(1.0) = \boxed{9.1 \times 10^{-6}\,\text{N}}. \tag{27.5.2}$$

That is about the weight of a single grain of sand, imperceptible on Earth. But in space it acts on the mirror every second, with nothing to oppose it. Give it a large enough sail and a small enough payload and the push becomes a voyage.

Example 27.6: Levitating a particle with light

A tiny black particle of mass $1.0\ \mu\text{g}$ absorbs all of a laser beam aimed straight up at it from below. What laser power would hold the particle up against gravity?

Solution

The force of the beam is its momentum delivered per second. A beam of power P delivers energy P each second, and hence, by equation (27.22), momentum P/c each second, so $F = P/c$. (Equivalently, $F = P_r A = (I/c)A = P/c$.) Setting this equal to the weight mg,

$$P = mgc = (1.0 \times 10^{-9}\,\text{kg})(9.8\,\text{m/s}^2)(3.00 \times 10^8\,\text{m/s}) \approx \boxed{2.9\,\text{W}}. \tag{27.5.3}$$

Three watts of light holds up a speck of dust a thousandth of a milligram. Radiation forces are feeble, yet at the microscopic scale they are large enough to use. Tightly focused laser beams, called *optical tweezers*, grab and move living cells and even single molecules, work honored with the 2018 Nobel Prize in Physics.

Checkpoint 27.5

Two identical satellites drift in sunlight, one painted matte black (a near-perfect absorber), the other mirror-bright (a near-perfect reflector). Which feels the larger radiation force, and by what factor?

Answer

Answers to Checkpoints

Checkpoint 27.5. The mirror-bright satellite, by a factor of 2. An absorbing surface takes up the light's momentum and so feels a pressure I/c; a reflecting surface reverses the momentum, so the change in momentum is twice as large and the pressure is $2I/c$. (It is the same reason a ball that bounces off a wall pushes it harder than one that sticks.)

This page titled 27.5: Momentum and Radiation Pressure of Electromagnetic Waves is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.6: The Electromagnetic Spectrum

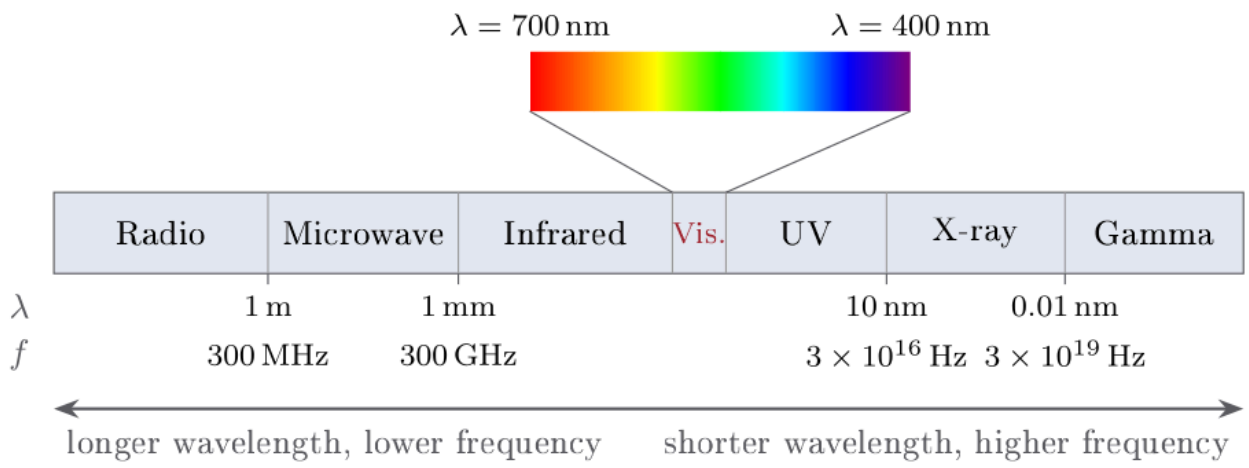


Figure 27.5. The electromagnetic spectrum. All electromagnetic waves travel at c in vacuum and obey $c = f\lambda$; they differ only in wavelength, from kilometer-scale radio waves to sub-atomic gamma rays. Visible light is the narrow band from roughly $400\ \text{nm}$ to $700\ \text{nm}$.

•Figure description

A horizontal bar showing the electromagnetic spectrum, divided left to right into Radio, Microwave, Infrared, a very narrow Visible band, Ultraviolet, X-ray and Gamma. Beneath the bar, the boundaries between bands are marked with their wavelengths and frequencies: 1 meter and 300 megahertz between radio and microwave, 1 millimeter and 300 gigahertz between microwave and infrared, 10 nanometers and 3 times 10 to the 16 hertz between ultraviolet and X-ray, and 0.01 nanometer and 3 times 10 to the 19 hertz between X-ray and gamma. A double-headed arrow along the bottom is labeled longer wavelength, lower frequency at the left end and shorter wavelength, higher frequency at the right end. Two lines lead from the thin visible band up to an enlarged strip showing the colors in order, red, orange, yellow, green, cyan, blue, violet, labeled wavelength 700 nanometers at the red end and 400 nanometers at the violet end. The visible strip is a tiny fraction of the whole bar.

Maxwell's equations place no restriction on the wavelength of an electromagnetic wave. Any wavelength is possible, and every one travels at the same speed c in vacuum, so that

$$c = f\lambda \tag{27.24}$$

ties the frequency to the wavelength. Sorted by wavelength (or, equivalently, by frequency), electromagnetic waves form the **electromagnetic spectrum** (Figure 27.5 and Table 27.1). The names of its regions are historical, and reflect how each kind of wave was discovered and is produced and detected, not any sharp physical boundary. They are all the same phenomenon, Maxwell's wave, differing only in wavelength and frequency.

Table 27.1. The regions of the electromagnetic spectrum. The boundaries are approximate and the regions overlap; X-rays and gamma rays, for instance, are distinguished more by how they are produced than by wavelength.

Region	Approximate wavelengths	Typical sources and uses
Radio	longer than $1\ \text{m}$	broadcasting, radio astronomy
Microwave	$1\ \text{mm}$ to $1\ \text{m}$	radar, cell phones, Wi-Fi, ovens
Infrared	$700\ \text{nm}$ to $1\ \text{mm}$	warm objects, remote controls
Visible	$400\ \text{nm}$ to $700\ \text{nm}$	the Sun, lamps, lasers; human vision
Ultraviolet	$10\ \text{nm}$ to $400\ \text{nm}$	the Sun, germicidal lamps; sunburn
X-ray	$0.01\ \text{nm}$ to $10\ \text{nm}$	X-ray tubes; medical imaging
Gamma ray	below about $0.01\ \text{nm}$	radioactive nuclei, cosmic sources

The range is enormous. The longest radio waves in use are kilometers long; the shortest gamma rays are smaller than an atomic nucleus. Yet the physics of their generation is the same at heart: *accelerating charges radiate*. Electrons surging back and forth in a transmitting antenna produce radio waves at the frequency of their oscillation. The jostling charged particles in any warm object produce infrared, and in a hot enough object visible light, the thermal radiation of Chapter 37. Electrons jumping between energy levels in atoms produce visible and ultraviolet light (Chapter 41); fast electrons slammed to a stop in a metal target produce X-rays; and nuclei rearranging themselves produce gamma rays (Chapter 43). Why are our eyes sensitive to this one narrow band? Several facts point to the same answer. The Sun's output peaks in the visible, and the atmosphere lets visible light through while absorbing most of the ultraviolet and much of the infrared. Water is even more selective: it is transparent to visible light but absorbs strongly at both longer and shorter wavelengths, and the first eyes evolved in the sea, long before animals moved onto land. Finally, visible light is energetic enough to drive the chemical change in the retina that starts the process of seeing, but not so energetic that it damages living tissue, as ultraviolet light does.

Example 27.7: Sizing a radio antenna

A simple radio antenna is a straight conducting rod, and it radiates or receives most efficiently when its length is half the wavelength, because then the oscillating current in it resonates. How long should such a half-wave antenna be for (a) an FM station at $98.5\ \text{MHz}$ and (b) a Wi-Fi router at $2.4\ \text{GHz}$?

Solution

(a) The wavelength is $\lambda = c/f = (3.00 \times 10^8)/(98.5 \times 10^6) = 3.05\ \text{m}$, so the antenna should be $\lambda/2 = \boxed{1.52\,\text{m}}$ long.

(b) Here $\lambda = (3.00 \times 10^8)/(2.4 \times 10^9) = 0.125\ \text{m}$, and $\lambda/2 = \boxed{6.3\,\text{cm}}$. The higher the frequency, the smaller the antenna. That is why the antennas inside a phone, which work at gigahertz frequencies, fit easily in a pocket, while AM radio towers, broadcasting near $1\ \text{MHz}$ with wavelengths of hundreds of meters, stand as tall as buildings.

Example 27.8: From wavelength to frequency

A laboratory lamp emits yellow light of wavelength $\lambda = 589\ \text{nm}$ (the sodium line). What is its frequency, and how does it compare with an FM radio station broadcasting at $100\ \text{MHz}$?

Solution

From equation (27.24),

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^8\,\text{m/s}}{589 \times 10^{-9}\,\text{m}} = \boxed{5.09 \times 10^{14}\,\text{Hz}}. \tag{27.6.1}$$

That is about $509{,}000{,}000\ \text{MHz}$, some five million times the frequency of the FM station. Both are the same kind of wave. The light simply oscillates millions of times faster and has a wavelength millions of times shorter (the FM wave's wavelength is $c/f = 3.0\ \text{m}$, about the length of a small car).

Checkpoint 27.6

Rank these by increasing frequency: a microwave-oven wave, a medical X-ray, green light, an AM radio wave. Then rank the same four by increasing wavelength. How are the two rankings related, and why?

Answer

Answers to Checkpoints

Checkpoint 27.6. By increasing frequency: AM radio wave, microwave-oven wave, green light, medical X-ray. By increasing wavelength the order is exactly reversed: X-ray, green light, microwave, AM radio wave. All electromagnetic waves travel at the same speed $c = f\lambda$ in vacuum, so frequency and wavelength are inversely proportional: the higher the frequency, the shorter the wavelength.

This page titled 27.6: The Electromagnetic Spectrum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.S: Electromagnetic Waves (Summary)

Summary

- A charging capacitor shows that Ampère's law is incomplete. The **displacement current** $i_d = \epsilon_0 \, d\Phi_E/dt$ repairs it: a changing electric field produces a magnetic field just as a current does.
- **Maxwell's equations**: two Gauss's laws, Faraday's law, and the Ampère–Maxwell law. Together with the Lorentz force they contain all of classical electromagnetism.
- In empty space, Faraday's law and the Ampère–Maxwell law combine into a wave equation. **Electromagnetic waves** are transverse: $\mathbf{E}$ and $\mathbf{B}$ are in phase, mutually perpendicular, and perpendicular to the travel direction $\mathbf{E} \times \mathbf{B}$, with $E/B = c$. They travel at $c = 1/\sqrt{\mu_0 \epsilon_0} \approx 3.00 \times 10^8 \text{ m/s}$.
- The **Poynting vector** $\mathbf{S} = \frac{1}{\mu_0}\mathbf{E} \times \mathbf{B}$ gives the instantaneous power per area. Its time average is the **intensity** $I = E_{\text{rms}}^2/(c\mu_0) = cB_{\text{rms}}^2/\mu_0 = E_m^2/(2c\mu_0)$. A point source obeys $I = P/4\pi r^2$.
- **Radiation pressure**: an electromagnetic wave carries momentum $p = U/c$, so it presses on surfaces: $P_r = I/c$ (absorbed) or $2I/c$ (reflected), with force $F = P_r A$.
- The **electromagnetic spectrum** runs from radio waves to gamma rays, all traveling at c with $c = f\lambda$. Visible light ($\sim$ 400–700 nm) is a thin band of it. Accelerating charges radiate.

Key Equations

Displacement current

$$i_d = \epsilon_0 \frac{d\Phi_E}{dt} \tag{27.S.1}$$

Maxwell's equations (integral form)

$$\begin{aligned}
\oint \mathbf{E}\cdot d\mathbf{A} &= \frac{q_{\text{enc}}}{\epsilon_0} && \text{(Gauss, electric)} \\
\oint \mathbf{B}\cdot d\mathbf{A} &= 0 && \text{(Gauss, magnetic)} \\
\oint \mathbf{E}\cdot d\mathbf{s} &= -\frac{d\Phi_B}{dt} && \text{(Faraday)} \\
\oint \mathbf{B}\cdot d\mathbf{s} &= \mu_0 i_{\text{enc}} + \mu_0\epsilon_0\frac{d\Phi_E}{dt} && \text{(Ampère–Maxwell)}
\end{aligned} \tag{27.S.2}$$

Speed of light in vacuum

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \approx 3.00 \times 10^8 \,\text{m/s}, \qquad c = f\lambda \tag{27.S.3}$$

Fields of an electromagnetic wave

$$\frac{E}{B} = c \tag{27.S.4}$$

Poynting vector

$$\mathbf{S} = \frac{1}{\mu_0}\,\mathbf{E} \times \mathbf{B}, \qquad S = \frac{EB}{\mu_0} = \frac{E^2}{c\mu_0} = \frac{cB^2}{\mu_0} \tag{27.S.5}$$

Intensity of an electromagnetic wave

$$I = S_{\text{avg}} = \frac{E_{\text{rms}}^2}{c\mu_0} = \frac{cB_{\text{rms}}^2}{\mu_0} = \frac{E_m^2}{2c\mu_0}, \qquad \text{point source: } I = \frac{P}{4\pi r^2} \tag{27.S.6}$$

Momentum of an electromagnetic wave

$$p = \frac{U}{c} \tag{27.S.7}$$

Radiation pressure

$$P_r = \frac{I}{c} \ \text{(absorbed)}, \qquad P_r = \frac{2I}{c} \ \text{(reflected)}, \qquad F = P_r A \tag{27.S.8}$$

This page titled 27.S: Electromagnetic Waves (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

27.E: Electromagnetic Waves (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $c = 3.00 \times 10^{8}\ \text{m/s}$, $\mu_0 = 1.257 \times 10^{-6}\ \text{T}\cdot\text{m/A}$, $\epsilon_0 = 8.85 \times 10^{-12}\ \text{C}^2/(\text{N}\cdot\text{m}^2)$, $c\mu_0 = 377\ \Omega$, and the solar intensity at Earth $I_{\text{sun}} = 1.36 \times 10^{3}\ \text{W/m}^2$.

Section 27.1 — Displacement Current

27.1 ★★ *[Conceptual.]* A parallel-plate capacitor is charging. (a) Is there a conduction current in the connecting wires? (b) Is there a conduction current in the gap between the plates? (c) What is present in the gap that produces a magnetic field there, and which term of the Ampère–Maxwell law accounts for it?

27.2 ★★ A parallel-plate capacitor has square plates $10\ \text{cm}$ on a side. The electric field between the plates is increasing at a rate of $1.5 \times 10^{12}\ \text{V/(m}\cdot\text{s)}$. Find the displacement current between the plates, and the conduction current in the wires.

27.3 ★★★ A capacitor with circular plates of radius R is charged by a current i. Show that the magnetic field between the plates is greatest at the edge of the plates, $r = R$, and find its value. Evaluate it for $R = 3.0\ \text{cm}$ and $i = 0.50\ \text{A}$.

Section 27.2 — Maxwell's Equations

27.4 ★ *[Conceptual.]* State in one sentence each what the four Maxwell equations physically assert. Which one would change if magnetic monopoles were discovered?

27.5 ★★ *[Conceptual.]* Faraday's law and the Ampère–Maxwell law each say that a changing field of one kind produces a circulating field of the other kind. Explain why this pair, rather than either one alone, is what makes electromagnetic waves possible.

Section 27.3 — Propagation of Electromagnetic Waves

27.6 ★ The electric field of a light wave has an amplitude of $250\ \text{V/m}$. Find the amplitude of its magnetic field.

27.7 ★★ *[Real-world application: astronomy.]* Light from the Sun takes about $8.3\ \text{min}$ to reach Earth. Use this to estimate the Earth–Sun distance.

27.8 ★★ Verify by direct substitution that $E = E_m \sin(kx - \omega t)$ satisfies equation (27.14) provided $\omega/k = 1/\sqrt{\mu_0\epsilon_0}$.

27.9 ★★ An electromagnetic wave travels in the $-z$ direction. At a certain point and instant its electric field points in the $+x$ direction. In which direction does its magnetic field point?

Section 27.4 — Energy Carried by Electromagnetic Waves

27.10 ★★ The electric field of a laser beam has an rms value of $E_{\text{rms}} = 500\ \text{V/m}$.

a. Find the rms magnetic field.
b. Find the intensity of the beam.

27.11 ★★ A $60\ \text{W}$ lightbulb radiates uniformly in all directions. Treating it as a point source, find the intensity of its light $2.0\ \text{m}$ away.

27.12 ★★★ *[Real-world application: engineering.]* A laser pointer emits $5.0\ \text{mW}$ in a beam of cross-sectional area $2.0\ \text{mm}^2$.

a. Find the intensity of the beam.
b. Find the rms electric and magnetic fields in the beam.

27.13 ★★ At a point where the electric-field amplitude of a radio wave is $E_m = 0.30\ \text{V/m}$, find (a) the magnetic-field amplitude and (b) the average intensity of the wave. (Recall $E = cB$ and $c\mu_0 = 377\ \Omega$.)

27.14 ★★ A cell-phone tower antenna radiates $20\ \text{W}$. Treating it as a point source radiating uniformly, at what distance is the intensity equal to $1.0 \times 10^{-3}\ \text{W/m}^2$? What is the rms electric field there?

Section 27.5 — Momentum and Radiation Pressure

27.15 ★★ Sunlight of intensity $1.36 \times 10^3\ \text{W/m}^2$ falls on a black (totally absorbing) solar panel of area $2.0\ \text{m}^2$ held face-on to the Sun. Find (a) the radiation pressure and (b) the radiation force on the panel.

27.16 ★★ Repeat the previous problem for a perfectly *reflecting* panel of the same area, and state how the force compares.

27.17 ★★★ *[Real-world application: spaceflight — the solar sail.]* Return to the spacecraft from the chapter opener. A solar sail of area $A = 32\ \text{m}^2$ is a perfect reflector, held face-on to sunlight at Earth's distance ($I_{\text{sun}} = 1.36 \times 10^3\ \text{W/m}^2$). The whole spacecraft has mass $m = 5.0\ \text{kg}$.

a. Find the radiation pressure on the sail.
b. Find the radiation force on the sail, and hence the acceleration of the spacecraft.
c. Starting from rest, how much speed does the sail gain in one day of continuous sunlight?
d. *[Conceptual.]* A chemical rocket can produce far more thrust than this. Why, then, might a solar sail still win a race to the outer solar system? (Think about fuel.)

27.18 ★★ *[Real-world application: optical tweezers.]* A laser beam of power $50\ \text{mW}$ is completely absorbed by a small particle. (a) Find the force on the particle. (b) What is the largest mass this beam could support against gravity? (c) Would the answer change if the particle reflected the beam straight back down?

Section 27.6 — The Electromagnetic Spectrum

27.19 ★ A radio station broadcasts at $91.5\ \text{MHz}$. Find the wavelength of its waves.

27.20 ★ Green light has a wavelength of $530\ \text{nm}$ in vacuum. Find its frequency.

27.21 ★★ A microwave oven uses radiation of frequency $2.45\ \text{GHz}$.

a. Find the wavelength of the microwaves.
b. The standing-wave pattern inside the oven creates hot spots half a wavelength apart. How far apart are they?

27.22 ★★ *[Conceptual.]* For each of the following, name the region of the electromagnetic spectrum and estimate its wavelength: (a) the signal from a Wi-Fi router at $5\ \text{GHz}$; (b) the radiation that makes a thermal camera image of a person; (c) the radiation that causes sunburn; (d) a dental X-ray.

27.23 ★ How long should a half-wave antenna be for a citizens-band radio operating at $27\ \text{MHz}$? For a $5.8\ \text{GHz}$ drone controller?

Challenge Problems

27.24 ★★★ *[Real-world application: astronomy — blowing dust out of the solar system.]* A spherical dust grain of radius a and density ρ sits a distance r from the Sun, which has luminosity $L_\odot = 3.85 \times 10^{26}\ \text{W}$ and mass $M_\odot = 1.99 \times 10^{30}\ \text{kg}$. The grain absorbs all the sunlight that falls on it.

a. Write the radiation force on the grain in terms of $L_\odot$, r, a, and c, and write the gravitational force on it in terms of $M_\odot$, r, a, and ρ.
b. Form the ratio $\beta = F_{\text{rad}}/F_{\text{grav}}$ and show that r cancels, so that whether a grain is blown out of the solar system depends only on its size and density, not on where it is.
c. Show that $\beta = 1$ at

$$a_{\text{crit}} = \frac{3L_\odot}{16\pi G M_\odot c\,\rho}, \tag{27.E.1}$$

and evaluate it for $\rho = 2500\ \text{kg/m}^3$. Express your answer in micrometers.

d. Grains much smaller than a_{crit} are expelled; grains much larger stay bound. Use this to explain why a comet's dust tail always points away from the Sun, and why the very smallest grains leave the fastest.

27.25 ★★★ *[Real-world application: radio.]* A radio station radiates $50\ \text{kW}$ uniformly in all directions (an idealization).

a. Find the intensity $20\ \text{km}$ away.
b. Find the amplitudes of the electric and magnetic fields there.
c. A receiving antenna is a straight wire $1.0\ \text{m}$ long, lined up with the electric field. Estimate the largest voltage induced along it.

This page titled 27.E: Electromagnetic Waves (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28: OSCILLATIONS

On September 19, 1985, a powerful earthquake struck off Mexico's Pacific coast, hundreds of kilometers from Mexico City. By the time the seismic waves reached the capital they should have been feeble, and on solid ground they were. But much of the city is built on the soft, water-soaked clay of an ancient lake bed, and that clay has a natural rhythm of its own: shaken, it sways back and forth about once every two seconds. The arriving waves happened to contain plenty of shaking at just that period, so the lake bed amplified them, rocking the city for minutes. Then came a second, cruel coincidence. Buildings, too, have natural periods of sway, and those roughly six to fifteen stories tall swayed at about the same two-second rhythm as the ground beneath them. Hundreds of them collapsed, while many shorter and taller buildings on the same streets came through with far less damage. What destroyed them was not simply the strength of the shaking but its *timing*: the ground pushed each building at very nearly the building's own natural rhythm, and a system driven at its natural rhythm responds with ever-growing amplitude. That runaway response has a name (**resonance**) and by the end of this chapter you will be able to say exactly what it is and why it is so dangerous.

A swaying building is an extreme example of something utterly ordinary. A mass bobbing on a spring, a child on a swing, the balance wheel of a wristwatch, a guitar string, the atoms in a crystal, the charge sloshing back and forth in the LC circuits of Chapter 26: all of these *oscillate*, repeating the same motion over and over about a stable equilibrium. Remarkably, an enormous number of these oscillators, despite their physical differences, obey the *same* equation and trace out the *same* motion in time. That shared motion is called simple harmonic motion, and it is the most important single model in all of physics. Master it here, for a mass on a spring, and you will meet it again in waves, in alternating-current circuits, in molecular vibrations, and in quantum mechanics.

Learning Objectives

- Recognize simple harmonic motion (SHM) as the motion produced by a linear restoring force, and write the position, velocity, and acceleration of an oscillator as functions of time.
- Relate the angular frequency of a mass–spring system to its mass and spring constant, and relate angular frequency, period, and frequency to one another.
- Compute the kinetic, potential, and total mechanical energy of an oscillator, and use energy conservation to find speeds and amplitudes.
- Show that a simple pendulum undergoes SHM for small angles, and find its period.
- Extend the analysis to torsion and physical pendulums.
- Describe how damping and an external driving force modify the motion, and explain the condition for resonance.

We will return to the Mexico City earthquake at the very end of the chapter, and in the final problems, once we have the tools to understand what went wrong.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

28.1: Simple Harmonic Motion
28.2: Hooke's Law
28.3: Mass–Spring System
28.4: Energy in Simple Harmonic Motion
28.5: The Simple Pendulum
28.6: Torsion Pendulum
28.7: Physical Pendulum

28.8: Damped Oscillations
28.9: Forced Oscillations
28.S: Oscillations (Summary)
28.E: Oscillations (Problems)

This page titled 28: Oscillations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.1: Simple Harmonic Motion

A motion that repeats itself at regular intervals is called **periodic**. The time for one complete cycle is the **period** T, measured in seconds, and the number of cycles per second is the **frequency** f, measured in hertz ($1\ \text{Hz} = 1\ \text{s}^{-1}$). The two are reciprocals,

Key Equation: Period and frequency

$$T = \frac{1}{f}, \tag{28.1}$$

and it will prove convenient to package the rate of oscillation in a third quantity, the **angular frequency** ω, which counts radians per second rather than cycles per second:

Key Equation: Angular frequency

$$\omega = \frac{2\pi}{T} = 2\pi f. \tag{28.2}$$

Because one full cycle corresponds to 2π radians, ω is larger than f by exactly that factor. Keeping ω, T, and f straight, and converting freely among them, is the small bit of bookkeeping that makes everything in this chapter run smoothly.

This page titled 28.1: Simple Harmonic Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.2: Hooke's Law

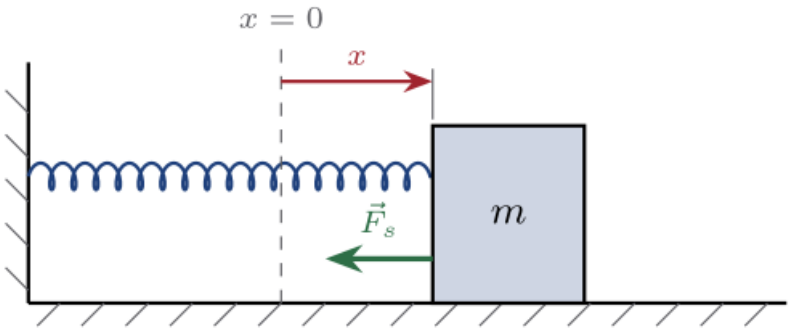


Figure 28.1. A block on a frictionless surface attached to a spring. Displaced a distance x to the right of equilibrium, the block feels a restoring force $\mathbf{F}_s = -k\mathbf{x}$ directed back toward $x = 0$. This linear restoring force is what produces simple harmonic motion.

•Figure description

A block of mass m on a horizontal frictionless floor, joined to a fixed wall on the left by a coiled spring. A dashed vertical line marks the equilibrium position x equals zero. The block has been pulled a distance x to the right of that line, and the stretched spring pulls it back with a restoring force F sub s pointing left, toward equilibrium.

What kind of force produces oscillation? Consider a block of mass m on a frictionless surface, attached to a spring whose other end is fixed (Figure 28.1). When the spring is at its natural length the block sits at its equilibrium position, which we take as $x = 0$. Pull the block to the right and the stretched spring pulls back to the left; push it to the left and the compressed spring pushes back to the right. In each case the spring force opposes the displacement: it is a **restoring force**, always directed back toward equilibrium. For an ideal spring its magnitude is proportional to the displacement, a relation known as **Hooke's law**:

Key Equation: Hooke's law

$$\mathbf{F}_s = -k\,\mathbf{x}. \tag{28.3}$$

The positive constant k is the **spring constant** (units $\mathrm{N/m}$); a stiff spring has a large k. The minus sign is the heart of the matter: it says the force points opposite to the displacement, and it is precisely this sign that makes the motion oscillate rather than run away. It takes twice the force to stretch the spring twice as far.

Simple harmonic motion (SHM) is, by definition, the motion of any system acted on by a restoring force proportional to the displacement, as in equation (28.3). The mass–spring system is the prototype, but as we will see the same mathematics governs the pendulum and a host of other systems.

This page titled 28.2: Hooke's Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.3: Mass–Spring System

To find the actual motion we combine Hooke's law with Newton's second law, $\mathbf{F}_{\text{net}} = m\mathbf{a}$. The only horizontal force on the block is the spring force, so

$$m\mathbf{a} = -k\mathbf{x} \qquad \Longrightarrow \qquad \mathbf{a} = -\frac{k}{m}\,\mathbf{x}. \tag{28.3.1}$$

The acceleration is proportional to the displacement and points in the opposite direction. Now recall that acceleration is the second derivative of position, $a = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}$. Substituting, and defining the constant $\omega^2 \equiv k/m$, we obtain the central equation of this chapter:

Key Equation: Equation of motion for SHM

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\omega^2 x, \qquad \omega = \sqrt{\frac{k}{m}}. \tag{28.4}$$

This is a **differential equation**: it relates a function to its own derivatives. It asks a precise question: what function $x(t)$ has the property that its second derivative is the function itself, times $-\omega^2$? The cosine and sine both have this property, and the general solution is

Key Equation: Position in SHM

$$x(t) = A\cos(\omega t + \phi). \tag{28.5}$$

Here A is the **amplitude**, the maximum displacement from equilibrium; ω is the angular frequency we just defined; and ϕ is the **phase constant**, fixed by where the block happens to be at $t = 0$. You should verify for yourself, by differentiating twice, that equation (28.5) indeed satisfies equation (28.4).

Since $T = 2\pi/\omega$ and $\omega = \sqrt{k/m}$, the period of a mass on a spring is

Key Equation: Period of a mass–spring system

$$T = 2\pi\sqrt{\frac{m}{k}}. \tag{28.6}$$

Notice the profound consequence: the period depends only on the stiffness k and the mass m, *not* on the amplitude. A block pulled far from equilibrium completes each cycle in exactly the same time as one barely displaced. This amplitude-independence is the signature of SHM and the reason a pendulum clock keeps good time even as its swing slowly dies away.

Velocity and acceleration

Differentiating the position once gives the velocity, and again gives the acceleration:

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = -A\omega\sin(\omega t + \phi), \tag{28.7}$$

$$[4pt]a(t) = \frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -A\omega^2\cos(\omega t + \phi). \tag{28.8}$$

Two facts about these expressions are worth committing to memory. First, the maximum speed and maximum acceleration are

$$v_{\max} = \omega A, \qquad a_{\max} = \omega^2 A, \tag{28.9}$$

reached respectively at the equilibrium point (where the cosine is zero, so the sine is largest) and at the turning points $x = \pm A$

(where the cosine is largest). The block is moving *fastest* as it whips through the middle and is momentarily *at rest* at the extremes, even though its acceleration is greatest there. Second, comparing equation (28.8) with equation (28.5) shows that

$$a(t) = -\omega^2 x(t), \tag{28.10}$$

which is just our original differential equation (28.4) restated, a reassuring sign that everything is consistent.

Checkpoint 28.1

A block oscillates on a spring in SHM. At the instant the block passes through the equilibrium position $x = 0$, which of the following are zero: its displacement, its velocity, its acceleration? At the instant it reaches a turning point $x = +A$, which are zero? *(Hint: read off equations (28.5)–(28.8) at the right phases.)*

Answer

Example 28.1: Timing the extremes of the motion

An object attached to a horizontal spring oscillates back and forth along a frictionless surface. Its maximum speed is $2.00\ \mathrm{m/s}$ and its maximum acceleration is $5.15\ \mathrm{m/s^2}$. How much time elapses between an instant when the object's speed is maximum and the next instant when its acceleration is maximum?

Solution

The speed is maximum at the equilibrium point and the acceleration is maximum at a turning point; the object travels from one to the other in a *quarter* of a full cycle. So the elapsed time we want is $T/4$, and the whole problem reduces to finding the period.

From the maxima in equation (28.9), $v_{\max} = \omega A$ and $a_{\max} = \omega^2 A$. Dividing the second by the first cancels the amplitude and leaves the angular frequency:

$$\frac{a_{\max}}{v_{\max}} = \frac{\omega^2 A}{\omega A} = \omega = \frac{5.15\,\mathrm{m/s^2}}{2.00\,\mathrm{m/s}} = 2.575\,\mathrm{rad/s}. \tag{28.3.2}$$

The period follows from equation (28.2):

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{2.575\,\mathrm{rad/s}} = 2.440\,\mathrm{s}, \tag{28.3.3}$$

so the time from maximum speed to the next maximum acceleration is

$$\frac{T}{4} = \frac{2.440\,\mathrm{s}}{4} \approx \boxed{0.610\,\mathrm{s}}. \tag{28.3.4}$$

As a check, note that we never needed A, k, or m individually: the ratio of the two maxima was enough, exactly because both share the same ω.

Answers to Checkpoints

Checkpoint 28.1. At $x = 0$ the displacement is zero and so is the acceleration, since $a = -\omega^2 x$; the velocity is at its maximum, ωA. At $x = +A$ the velocity is zero (the block is momentarily at rest as it turns around) while the acceleration has its maximum size, $\omega^2 A$, directed back toward equilibrium. Displacement and acceleration are zero together, and the velocity is zero exactly when they are largest.

This page titled 28.3: Mass–Spring System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.4: Energy in Simple Harmonic Motion

As the block oscillates, energy sloshes back and forth between two forms. When the block races through equilibrium it has no spring stretch but maximum speed: the energy is all kinetic. When it pauses at a turning point it has zero speed but maximum stretch: the energy is all potential. Everywhere in between, it is some of each. Let us make this quantitative.

The **kinetic energy** is $K = \frac{1}{2}mv^2$. Using the velocity from equation (28.7),

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}mA^2\omega^2 \sin^2(\omega t + \phi). \tag{28.11}$$

The **elastic potential energy** stored in the spring is $U_s = \frac{1}{2}kx^2$. Using the position from equation (28.5),

$$U_s = \tfrac{1}{2}kx^2 = \tfrac{1}{2}kA^2 \cos^2(\omega t + \phi). \tag{28.12}$$

Both expressions involve a squared trigonometric function, so both are always positive or zero, as energies must be. Now add them to get the total mechanical energy, and use the defining relation $\omega^2 = k/m$ (so that $m\omega^2 = k$) to put the kinetic term in the same form as the potential:

$$E = K + U_s = \tfrac{1}{2}kA^2 \sin^2(\omega t + \phi) + \tfrac{1}{2}kA^2 \cos^2(\omega t + \phi) = \tfrac{1}{2}kA^2\left[\sin^2(\omega t + \phi) + \cos^2(\omega t + \phi)\right]. \tag{28.4.1}$$

The bracket is the Pythagorean identity, equal to 1 at every instant. The time dependence cancels completely, leaving a constant:

Key Equation: Total energy of a simple harmonic oscillator

$$E = \tfrac{1}{2}kA^2. \tag{28.13}$$

Conservation of Energy

The total mechanical energy of an undamped oscillator is constant in time and is fixed entirely by the amplitude. This is conservation of energy at work: with no friction to drain it, energy merely shifts between kinetic and potential, and

$$K_1 + U_1 = K_2 + U_2 \tag{28.14}$$

for any two instants 1 and 2. Because $E = \frac{1}{2}kA^2$ is also equal to the kinetic energy at equilibrium, $\frac{1}{2}mv_{\max}^2$, we have a quick route to the maximum speed: $v_{\max} = A\sqrt{k/m} = \omega A$, as we found before.

Example 28.2: A skyscraper's mass damper

Many tall buildings contain a *tuned mass damper*: a massive block mounted on springs near the top of the tower, free to slide on a lubricated track, that sways out of step with the building to cancel its motion in a high wind or earthquake. Suppose such a block has mass $m = 2.72 \times 10^5\ \text{kg}$ and is designed to oscillate in SHM at frequency $f = 0.150\ \text{Hz}$ (matching the slow sway of the tower itself) with amplitude $A = 1.00\ \text{m}$.

a. What is the total mechanical energy of the spring–block system?
b. What is the block's speed as it passes through the equilibrium point?

Solution

First find the angular frequency and the effective spring constant. From equation (28.2),

$$\omega = 2\pi f = 2\pi(0.150\,\text{Hz}) = 0.9425\,\text{rad/s}, \tag{28.4.2}$$

and since $\omega = \sqrt{k/m}$, the spring constant is $k = m\omega^2 = (2.72 \times 10^5)(0.9425)^2 = 2.416 \times 10^5\ \text{N/m}$.

(a) With $A = 1.00\ \text{m}$, the total energy is

$$E = \tfrac{1}{2}kA^2 = \tfrac{1}{2}(2.416\times10^5\,\mathrm{N/m})(1.00\,\mathrm{m})^2 \approx \boxed{1.21\times10^5\,\mathrm{J}}. \tag{28.4.3}$$

(b) At equilibrium the energy is entirely kinetic, $E = \tfrac{1}{2}mv_{\max}^2$, so

$$v_{\max} = \sqrt{\frac{2E}{m}} = \sqrt{\frac{2(1.21\times10^5\,\mathrm{J})}{2.72\times10^5\,\mathrm{kg}}} \approx \boxed{0.942\,\mathrm{m/s}}. \tag{28.4.4}$$

As a check, the same speed comes directly from $v_{\max} = \omega A = (0.9425)(1.00) = 0.942\ \mathrm{m/s}$: about $3.4\ \mathrm{km/hr}$, a slow walking pace, but for a block weighing as much as a fully loaded airliner. (A damper must be tuned to the building's own natural frequency, which for a skyscraper is well below $1\ \mathrm{Hz}$; the damper atop Taipei 101, for example, sways with a period of about $7\ \mathrm{s}$.)

Example 28.3: A vibrating diving board

A $76.6\ \mathrm{kg}$ diver stands at the end of a diving board while it vibrates up and down in SHM. The board has an effective spring constant $k = 3100\ \mathrm{N/m}$, and the vertical distance between the highest and lowest points of the motion is $0.214\ \mathrm{m}$.

a. What is the amplitude of the motion?
b. Starting from the highest point, what is the diver's speed one-quarter of a period later?
c. If the distance between highest and lowest points were instead $0.141\ \mathrm{m}$, what would be the time for one complete cycle?

Solution

(a) The amplitude is the maximum displacement from equilibrium, which is *half* the full peak-to-peak distance:

$$A = \tfrac{1}{2}(0.214\,\mathrm{m}) = \boxed{0.107\,\mathrm{m}}. \tag{28.4.5}$$

(b) The angular frequency of the diver-on-board oscillator is

$$\omega = \sqrt{\frac{k}{m}} = \sqrt{\frac{3100\,\mathrm{N/m}}{76.6\,\mathrm{kg}}} = 6.362\,\mathrm{rad/s}. \tag{28.4.6}$$

Starting at the highest point (a turning point, where the speed is zero) and advancing a quarter period brings the diver to the equilibrium point, where the speed is maximum:

$$v_{\max} = \omega A = (6.362\,\mathrm{rad/s})(0.107\,\mathrm{m}) \approx \boxed{0.681\,\mathrm{m/s}}. \tag{28.4.7}$$

(c) Here is the trap. The period of SHM depends on k and m but *not* on the amplitude, so changing the peak-to-peak distance changes nothing about the period:

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{76.6\,\mathrm{kg}}{3100\,\mathrm{N/m}}} \approx \boxed{0.988\,\mathrm{s}}. \tag{28.4.8}$$

The smaller amplitude means the diver moves more slowly and over a shorter distance, in just such proportion that each cycle still takes the same $0.988\ \mathrm{s}$.

Checkpoint 28.2

A block oscillates on a spring with amplitude A. (a) At what displacement are its kinetic and potential energies equal? (b) If the amplitude is doubled, by what factor do the total energy, the maximum speed, and the period change?

Answer

Answers to Checkpoints

Checkpoint 28.2. (a) The total energy is $\frac{1}{2}kA^2$. Equal shares means $U = \frac{1}{2}kx^2$ is half of that: $x^2 = A^2/2$, so $x = \pm A/\sqrt{2} \approx \pm 0.71A$. Not halfway out, because potential energy grows as x^2. (b) The total energy, $\frac{1}{2}kA^2$, quadruples. The maximum speed, $v_{\max} = \omega A$, doubles. The period, $2\pi\sqrt{m/k}$, does not change at all: it is independent of the amplitude.

This page titled 28.4: Energy in Simple Harmonic Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.5: The Simple Pendulum

A **simple pendulum** is a single point mass (the *bob*) hanging from a string of fixed length ℓ, free to swing back and forth. There is more than one kind of pendulum, as we will see, but this idealized one (all the mass at a point, the string massless and inextensible) is the place to start. Does it undergo SHM?

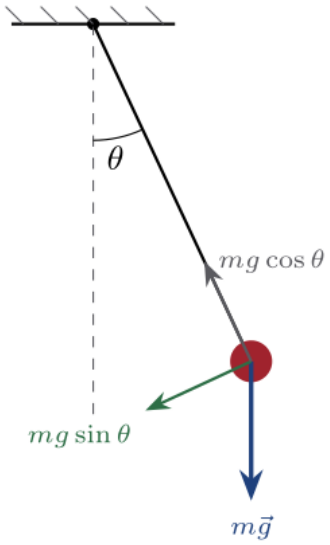


Figure 28.2. A simple pendulum of length ℓ displaced by angle θ. Gravity's component perpendicular to the string, $mg\sin\theta$, is the restoring force. For small angles $\sin\theta \approx \theta$, and the motion becomes simple harmonic.

•Figure description

A simple pendulum hanging from a fixed pivot at the ceiling. A dashed vertical line drops straight down from the pivot, and the string holding the bob makes an angle theta with it. Three arrows are drawn at the bob: the full weight m g pointing straight down, its component m g sine theta perpendicular to the string, which is the restoring force pulling the bob back toward the lowest point, and its component m g cosine theta along the string, away from the pivot.

When the string makes an angle θ with the vertical (Figure 28.2), two forces act on the bob: the tension along the string and gravity, mg, straight down. Resolve gravity into a component along the string (balanced by the tension) and a component perpendicular to the string. Only the perpendicular component is unbalanced, and it is what accelerates the bob along its arc:

$$F_\perp = -mg\sin\theta. \tag{28.5.1}$$

The minus sign marks it as a restoring force: whichever way the bob swings, this component pulls it back toward the bottom. But notice the difficulty: the force depends on $\sin\theta$, not on θ itself, so it is *not* proportional to the displacement, and the motion is not exactly SHM.

Here we make the crucial **small-angle approximation**. For small angles (measured in radians), $\sin\theta \approx \theta$, so the restoring force becomes

$$F_\perp \approx -mg\,\theta, \tag{28.5.2}$$

which *is* proportional to the (angular) displacement. With this approximation the pendulum becomes a simple harmonic oscillator. Writing the bob's displacement along its arc as $s = \ell\theta$ and applying Newton's second law along the arc, $F_\perp = m\,\frac{\mathrm{d}^2 s}{\mathrm{d}t^2}$, gives

$$-mg\theta = m\ell\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} \qquad \Longrightarrow \qquad \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} = -\frac{g}{\ell}\,\theta. \tag{28.5.3}$$

Compare this with the master equation (28.4): it has exactly the same form, with the angle θ playing the role of x and g/ℓ playing the role of ω^2. The solution is therefore the familiar $\theta(t) = \theta_{\max}\cos(\omega t + \phi)$, with angular frequency $\omega = \sqrt{g/\ell}$ and hence a period

Key Equation: Period of a simple pendulum

$$T = 2\pi\sqrt{\frac{\ell}{g}}. \tag{28.15}$$

The result is striking: the period of a simple pendulum depends *only* on its length and the local value of g. It does not depend on the mass of the bob, and (within the small-angle approximation) it does not depend on the amplitude of the swing. This is why a pendulum makes such a good clock, and it is what you confirmed experimentally in your first Physics 4A laboratory.

Checkpoint 28.3

Two pendulums have bobs of different mass but the same length, and swing through the same small angle. Which has the longer period, or are they equal? Now suppose instead the two have the same mass but different lengths. Which has the longer period? Answer using equation (28.15), not intuition alone.

Answer

Example 28.4: Measuring a tower with a pendulum

A rope hangs down the center of a spiral staircase in an old castle tower, fixed at the top and reaching nearly to the floor. With nothing on hand to measure its length, you tie a small object to the bottom end to make a simple pendulum that just clears the floor, set it swinging through a small angle, and time its period at $6.17\ \text{s}$. How tall is the tower?

Solution

The pendulum's length is essentially the height of the tower, so finding ℓ answers the question. Solve equation (28.15) for ℓ:

$$T = 2\pi\sqrt{\frac{\ell}{g}} \quad \Longrightarrow \quad \ell = g\left(\frac{T}{2\pi}\right)^2. \tag{28.5.4}$$

Substituting $T = 6.17\ \text{s}$ and $g = 9.8\ \text{m/s}^2$,

$$\ell = (9.8\,\text{m/s}^2)\left(\frac{6.17\,\text{s}}{2\pi}\right)^2 = (9.8)(0.9820)^2 \approx \boxed{9.45\,\text{m}}. \tag{28.5.5}$$

About three stories: a reasonable height for a castle tower, and a measurement made with nothing more than a stopwatch and a rock.

Answers to Checkpoints

Checkpoint 28.3. Equal. For small angles, $T = 2\pi\sqrt{\ell/g}$ contains no mass: a heavier bob has more weight pulling it back, but proportionally more inertia to move, and the two effects cancel. With the same mass and different lengths, the *longer* pendulum has the longer period, since $T \propto \sqrt{\ell}$; a pendulum four times as long takes twice as long to swing.

This page titled 28.5: The Simple Pendulum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.6: Torsion Pendulum

The simple pendulum is not the only oscillator built on a restoring torque. Two close relatives appear often enough to deserve their own formulas: the torsion pendulum of this section and the physical pendulum of the next.

Suspend a disk or a bar from a thin wire fixed at its center and give it a twist. The wire resists, exerting a restoring *torque* proportional to the angle of twist, $\tau = -\kappa\theta$, where κ (the Greek letter kappa) is the **torsion constant** of the wire. This is Hooke's law for twisting, and the rotational form of Newton's second law, $\tau = I\,\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}$, leads once again to an SHM equation: now with the moment of inertia I in place of mass and κ in place of k. The period is

Key Equation: Period of a torsion pendulum

$$T = 2\pi\sqrt{\frac{I}{\kappa}}. \tag{28.16}$$

The torsion pendulum is the working principle of the balance wheel in a mechanical watch and of the ultra-sensitive instruments used to measure tiny forces, including Cavendish's famous weighing of the Earth.

This page titled 28.6: Torsion Pendulum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.7: Physical Pendulum

A **physical pendulum** is any rigid body that swings about a fixed pivot under gravity: a swinging baseball bat, a leg, a hanging signboard. Unlike the idealized simple pendulum, its mass is spread out rather than concentrated at a point. Let I be the body's moment of inertia about the pivot and ℓ_{cm} the distance from the pivot to the center of mass. The restoring torque for small angles is $\tau = -mg\ \ell_{\mathrm{cm}}\ \theta$, and the same rotational analysis gives the period

Key Equation: Period of a physical pendulum

$$T = 2\pi\sqrt{\frac{I}{mg\,\ell_{\mathrm{cm}}}}. \tag{28.17}$$

This formula contains the simple pendulum as a special case. If all the mass sits at a single point a distance ℓ from the pivot, then $I = m\ell^2$ and $\ell_{\mathrm{cm}} = \ell$, so

$$T = 2\pi\sqrt{\frac{m\ell^2}{mg\ell}} = 2\pi\sqrt{\frac{\ell}{g}}, \tag{28.7.1}$$

recovering equation (28.15) exactly, a good check that the more general result is sensible.

This page titled 28.7: Physical Pendulum is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.8: Damped Oscillations

Every real oscillator eventually stops. A struck tuning fork fades to silence; a child on a swing coasts to rest if no one pushes. The reason is **damping**: a resistive force, usually from friction or from motion through air or fluid, that opposes the velocity and drains mechanical energy. For slow motion this drag force is well modeled as proportional to the speed, $F_{\text{damp}} = -b\ \frac{\mathrm{d}x}{\mathrm{d}t}$, where b is a positive **damping constant**.

Adding this force to the spring force changes the equation of motion. Newton's second law now reads $m\ \frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -kx - b\ \frac{\mathrm{d}x}{\mathrm{d}t}$, which we rearrange into the standard form

Key Equation: Equation of motion for a damped oscillator

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + kx = 0. \tag{28.18}$$

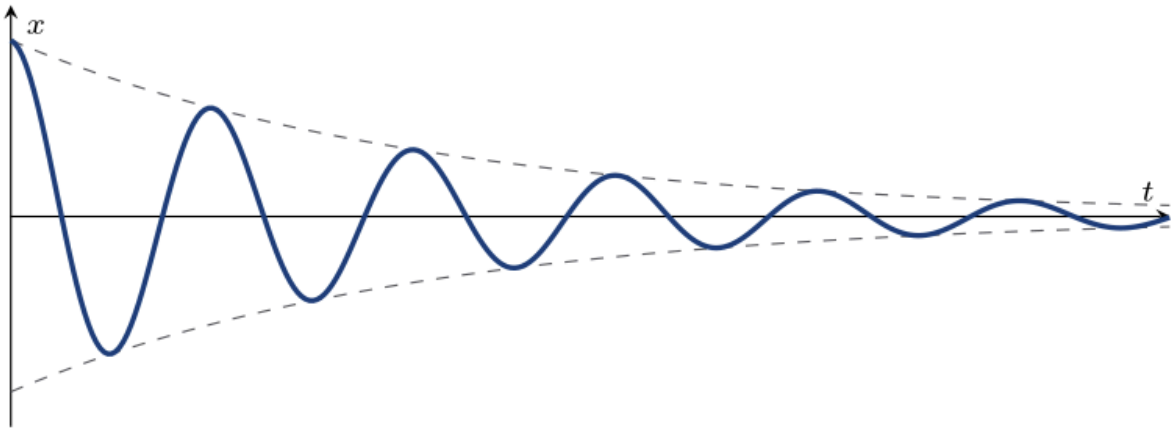


Figure 28.3. Underdamped motion. The oscillation continues but its amplitude decays under the exponential envelope $\pm A\ e^{-(b/2m)t}$ (dashed). With stronger damping the system returns to rest without overshooting (*critically damped*) or sluggishly without oscillating at all (*overdamped*).

•Figure description

A graph of displacement x against time t for an underdamped oscillator. The curve oscillates back and forth about zero while its swings get steadily smaller, and it stays inside a pair of dashed curves that decay exponentially toward the time axis, one above and one below. Those dashed curves are the envelope plus and minus A times e to the power minus b over 2 m times t.

For light damping the solution is an oscillation whose amplitude decays exponentially in time:

$$x(t) = A\,e^{-(b/2m)\,t}\cos(\omega' t + \phi). \tag{28.19}$$

The cosine still carries the back-and-forth, but the once-constant amplitude A is now multiplied by the **damping factor** $e^{-(b/2m)t}$, an envelope that shrinks the swing cycle by cycle (Figure 28.3). The angular frequency ω' is slightly less than the undamped value $\sqrt{k/m}$, since the drag also slows the oscillation a little.

How quickly the motion dies depends on the size of b, and three regimes are distinguished. With light damping the system is **underdamped**: it oscillates many times within a slowly shrinking envelope, as in Figure 28.3. Increase the damping and the oscillations disappear: a **critically damped** system returns to equilibrium in the shortest possible time without overshooting, the behavior engineers design into a car's shock absorbers and a door closer. Increase it further still and the system is **overdamped**, creeping back to equilibrium even more slowly, as though moving through molasses.

Checkpoint 28.4

You are designing the suspension of a car. Would you want it underdamped, critically damped, or overdamped? Explain what the ride would feel like in each case, and which gives the smoothest return to level after hitting a bump.

Answer

Answers to Checkpoints

Checkpoint 28.4. Critically damped (or just slightly underdamped) is the goal. An underdamped suspension lets the car bounce up and down several times after each bump: a floaty, seasick ride. An overdamped one is stiff: it transmits more of each jolt to the passengers and then creeps slowly back to level. Critical damping returns the car to level in the shortest time without overshooting, the smoothest return. Real cars are tuned slightly underdamped, trading a small overshoot for a quicker response.

This page titled 28.8: Damped Oscillations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.9: Forced Oscillations

Left alone, every oscillator vibrates at its own **natural frequency**, set by its physical makeup. For a mass–spring system,

$$\omega_0 = \sqrt{\frac{k}{m}}. \tag{28.20}$$

But we can also drive an oscillator with an external periodic force, pushing it at some **driving frequency** ω_d of our choosing. The classic picture is a hand moving the top of a spring up and down: the hand moves at ω_d, and after a brief transient the block settles into steady oscillation at that same driving frequency, but with an amplitude that depends dramatically on how ω_d compares with the natural ω_0.

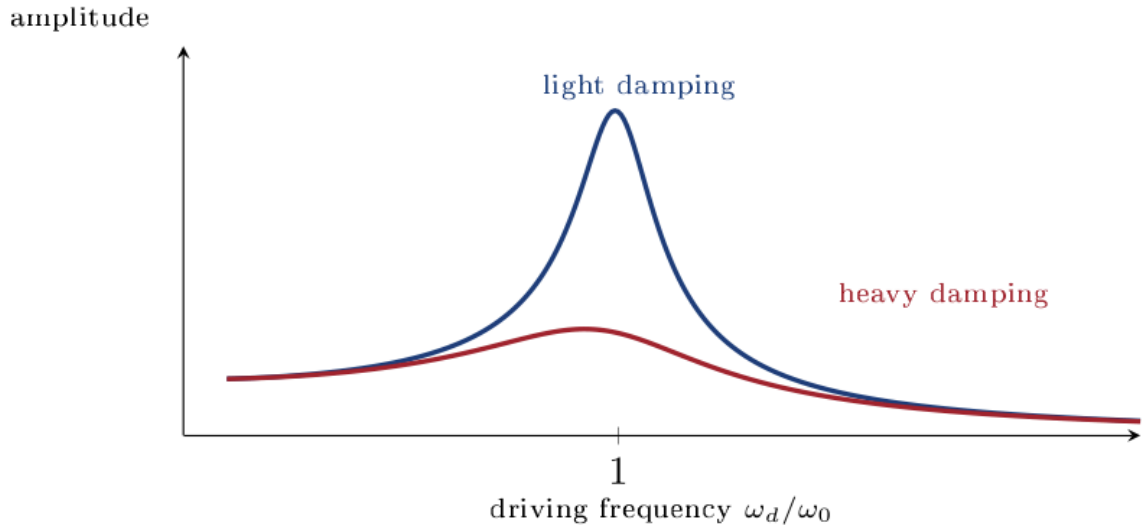


Figure 28.4. Resonance. The steady-state amplitude of a driven oscillator peaks sharply when the driving frequency equals the natural frequency, $\omega_d = \omega_0$. The lighter the damping, the taller and narrower the peak.

•Figure description

A graph of steady-state amplitude against the ratio of driving frequency to natural frequency, omega sub d over omega sub zero. Two curves are shown. The light-damping curve rises to a tall, narrow peak at a ratio of one and falls away steeply on either side. The heavy-damping curve peaks at the same place but is much lower and far broader. Both approach zero at high frequency ratios.

When the driving frequency is far from ω_0, the block responds weakly. As ω_d approaches ω_0, the response grows. And when $\omega_d = \omega_0$ (when we push exactly in step with the system's own rhythm, adding energy on every cycle at just the right moment) the amplitude grows enormously. This dramatic build-up is **resonance**. The only thing that limits the amplitude at resonance is damping; with little damping, the response at resonance can become large enough to destroy the system (Figure 28.4).

Resonance is everywhere. It is how a radio tunes to one station and rejects the rest, how a quartz crystal keeps a watch on time, how a pushed swing climbs higher with each well-timed push, and how an opera singer can shatter a glass by sounding its natural note.

Resonance also has a long record of causing engineering trouble. In 1831 a column of soldiers marching in step across the Broughton Suspension Bridge near Manchester felt it begin to vibrate in time with their footsteps; the extra loading snapped a bolt in one of its chains and the bridge collapsed beneath them. Ever since, the British Army has ordered troops to break step when crossing a bridge. In June 2000, London's new Millennium Bridge began swaying from side to side on its opening weekend. Its sideways natural frequency, about $1\ \mathrm{Hz}$, was close to the rate at which a walker's weight shifts from foot to foot; once the deck started to move, pedestrians unconsciously fell into step with the sway, pushing it harder still. The bridge closed after two days and reopened in 2002 fitted with dampers. And in the Mexico City earthquake of the chapter opener, the lake-bed clay drove the city's mid-rise buildings at their own natural period. The engineering responses share one idea: keep a structure's natural

frequencies away from the frequencies likely to drive it, and build in enough damping that no resonance can grow without bound, the same reason a skyscraper carries the tuned mass damper of our earlier example.

Checkpoint 28.5

Two identical oscillators, one lightly damped and one heavily damped, are driven by the same periodic force. (a) Driven at the natural frequency, which one reaches the larger amplitude? (b) Driven at a frequency far below the natural frequency, does the damping matter much? Explain using Figure 28.4.

Answer

Answers to Checkpoints

Checkpoint 28.5. (a) The lightly damped one, by far. At resonance each push arrives in step with the motion and adds energy, and the steady-state amplitude grows until the energy lost to damping per cycle balances the energy supplied; with little damping that balance comes only at a large amplitude, giving the tall, narrow peak of Figure 28.4. (b) Hardly at all. Far below resonance the oscillator simply follows the slowly varying force, its motion set mainly by the spring's stiffness, and the two curves in the figure nearly coincide. Damping matters most near resonance.

This page titled 28.9: Forced Oscillations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.S: Oscillations (Summary)

Summary

- **Periodic** motion repeats with period T and frequency $f = 1/T$. The **angular frequency** is $\omega = 2\pi/T = 2\pi f$.
- **Simple harmonic motion** (SHM) is the motion produced by a linear restoring force, $\mathbf{F}_s = -k\mathbf{x}$ (**Hooke's law**). It obeys $\mathrm{d}^2x/\mathrm{d}t^2 = -\omega^2 x$, with solution $x(t) = A\cos(\omega t + \phi)$.
- For a mass–spring system $\omega = \sqrt{k/m}$, so the period $T = 2\pi\sqrt{m/k}$ is *independent of amplitude*. The velocity is $v = -A\omega\sin(\omega t + \phi)$ and acceleration $a = -A\omega^2\cos(\omega t + \phi) = -\omega^2 x$; the maxima are $v_{\max} = \omega A$ and $a_{\max} = \omega^2 A$.
- **Energy**: $K = \frac{1}{2}mv^2$ and $U_s = \frac{1}{2}kx^2$ trade off, while the total $E = \frac{1}{2}kA^2$ stays constant. Energy conservation, $K_1 + U_1 = K_2 + U_2$, gives speeds and amplitudes quickly.
- A **simple pendulum** undergoes SHM for small angles ($\sin\theta \approx \theta$), with period $T = 2\pi\sqrt{\ell/g}$, independent of mass and amplitude.
- Related oscillators: the **torsion pendulum**, $T = 2\pi\sqrt{I/\kappa}$, and the **physical pendulum**, $T = 2\pi\sqrt{I/(mg\,\ell_{\mathrm{cm}})}$, which reduces to the simple pendulum when $I = m\ell^2$.
- **Damping** (a force $-b\ \mathrm{d}x/\mathrm{d}t$) makes the amplitude decay as $e^{-(b/2m)t}$; systems are *underdamped*, *critically damped*, or *overdamped*.
- A **driven** oscillator settles at the driving frequency; its amplitude peaks sharply at **resonance**, $\omega_d = \omega_0 = \sqrt{k/m}$, limited only by damping.

Key Equations

Period, frequency, and angular frequency

$$T = \frac{1}{f}, \qquad \omega = \frac{2\pi}{T} = 2\pi f \tag{28.S.1}$$

Hooke's law

$$\mathbf{F}_s = -k\,\mathbf{x} \tag{28.S.2}$$

Equation of motion for SHM

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\omega^2 x, \qquad \omega = \sqrt{\frac{k}{m}} \tag{28.S.3}$$

Position, velocity, and acceleration in SHM

$$\begin{aligned} x(t) &= A\cos(\omega t + \phi) \\ v(t) &= -A\omega\sin(\omega t + \phi) \\ a(t) &= -A\omega^2\cos(\omega t + \phi) = -\omega^2 x \end{aligned} \tag{28.S.4}$$

Maximum speed and acceleration

$$v_{\max} = \omega A, \qquad a_{\max} = \omega^2 A \tag{28.S.5}$$

Period of a mass–spring system

$$T = 2\pi\sqrt{\frac{m}{k}} \tag{28.S.6}$$

Energy of a simple harmonic oscillator

$$E = K + U_s = \tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = \tfrac{1}{2}kA^2 \tag{28.S.7}$$

Period of a simple pendulum

$$T = 2\pi\sqrt{\frac{\ell}{g}} \tag{28.S.8}$$

Period of a torsion pendulum

$$T = 2\pi\sqrt{\frac{I}{\kappa}} \tag{28.S.9}$$

Period of a physical pendulum

$$T = 2\pi\sqrt{\frac{I}{mg\,\ell_{\text{cm}}}} \tag{28.S.10}$$

Equation of motion for a damped oscillator

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + b\frac{\mathrm{d}x}{\mathrm{d}t} + kx = 0 \tag{28.S.11}$$

Damped amplitude envelope

$$x(t) = A\,e^{-bt/2m}\cos(\omega' t + \phi) \tag{28.S.12}$$

Resonance condition

$$\omega_d = \omega_0 = \sqrt{\frac{k}{m}} \tag{28.S.13}$$

This page titled 28.S: Oscillations (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

28.E: Oscillations (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Take $g = 9.8\ \text{m/s}^2$ throughout, and assume small-angle, frictionless, ideal conditions unless told otherwise.

Sections 28.1–28.2 — Simple Harmonic Motion and Hooke's Law

28.1 ★ A mass on a spring completes 20 full oscillations in $8.0\ \text{s}$.

a. What is the period of the motion?
b. What is the frequency?
c. What is the angular frequency?

28.2 ★ A spring stretches $4.0\ \text{cm}$ when a $2.0\ \text{kg}$ mass is hung from it at rest. What is the spring constant k?

28.3 ★★ *[Conceptual.]* A block oscillating on a spring is at its equilibrium position at one instant and at its maximum displacement a short time later.

a. At which of these two points is the speed greatest? At which is the magnitude of the acceleration greatest?
b. Explain why the block is momentarily at rest exactly where its acceleration is largest.

Section 28.3 — Mass–Spring System

28.4 ★ A $0.50\ \text{kg}$ block oscillates on a spring of constant $k = 200\ \text{N/m}$.

a. Find the angular frequency ω.
b. Find the period T and frequency f.

28.5 ★★ A $0.25\ \text{kg}$ mass on a spring moves as $x(t) = (0.080\ \text{m})\cos\!\left[(12\ \text{rad/s})\ t\right]$.

a. State the amplitude, angular frequency, and period.
b. Find the maximum speed and maximum acceleration.
c. Find the spring constant.

28.6 ★★ *[Real-world application: engineering.]* The maximum speed of a piston in SHM is $8.0\ \text{m/s}$ and its maximum acceleration is $640\ \text{m/s}^2$. Find (a) the angular frequency, (b) the amplitude, and (c) the period.

28.7 ★★ A $0.25\ \text{kg}$ mass on a spring oscillates with a period of $0.80\ \text{s}$.

a. Find the spring constant.
b. If the amplitude is $6.0\ \text{cm}$, find the maximum speed and the maximum acceleration of the mass.

Section 28.4 — Energy in Simple Harmonic Motion

28.8 ★ A spring of constant $k = 350\ \text{N/m}$ oscillates with amplitude $6.0\ \text{cm}$. What is the total mechanical energy of the system?

28.9 ★★ A $0.40\ \text{kg}$ block on a spring ($k = 160\ \text{N/m}$) oscillates with amplitude $0.12\ \text{m}$.

a. Find the total energy.
b. Find the maximum speed.
c. Find the speed when the block is $0.060\ \text{m}$ from equilibrium.

28.10 ★★ *[Real-world application: biology.]* A web-building spider of mass $0.50\ \text{g}$ sits at the center of its web, which acts as a spring of constant $k = 0.30\ \text{N/m}$, and bounces in SHM with amplitude $3.0\ \text{mm}$. Find (a) the total energy of the oscillation and (b)

the spider's maximum speed.

Section 28.5 — The Simple Pendulum

28.11 ★ A simple pendulum has length $1.2\ \text{m}$. What is its period for small-amplitude swings?

28.12 ★ You want a simple pendulum whose period is exactly $2.0\ \text{s}$ (a "seconds pendulum," which ticks once per second). How long should it be?

28.13 ★★ *[Conceptual.]* A grandfather clock keeps time with a pendulum. On a hot day the metal rod expands slightly, lengthening the pendulum.

a. Does the clock now run fast or slow? Justify your answer with equation (28.15).
b. A clockmaker corrects it by adjusting the bob's position. Which way should the bob be moved?

Sections 28.6–28.7 — Torsion and Physical Pendulums

28.14 ★★ A uniform rod of mass $0.30\ \text{kg}$ and length $0.80\ \text{m}$ swings as a physical pendulum about one end. Its moment of inertia about that end is $I = \frac{1}{3}m\ell^2$, and its center of mass is at $\ell_{\text{cm}} = \ell/2$. Find the period of small oscillations.

28.15 ★★ A torsion pendulum consists of a disk of moment of inertia $I = 2.5 \times 10^{-3}\ \text{kg}\cdot\text{m}^2$ hanging from a wire of torsion constant $\kappa = 0.040\ \text{N}\cdot\text{m/rad}$. Find its period.

Section 28.8 — Damped Oscillations

28.16 ★★ A damped oscillator has mass $m = 0.25\ \text{kg}$ and damping constant $b = 0.070\ \text{kg/s}$. Its amplitude decays as $A e^{-(b/2m)t}$.

a. How long does it take the amplitude to fall to half its initial value?
b. *[Conceptual.]* Does the period of the (slow) oscillation change as the amplitude shrinks? Explain.

28.17 ★ *[Conceptual.]* Classify each as underdamped, critically damped, or overdamped, and justify briefly: (a) a screen door with a good pneumatic closer that shuts smoothly without slamming; (b) a child's swing given one push and left alone; (c) a heavy gate moving slowly through thick grease that never quite oscillates.

Section 28.9 — Forced Oscillations & the Chapter Opener

28.18 ★ A mass–spring system has $m = 2.0\ \text{kg}$ and $k = 50\ \text{N/m}$. At what driving frequency ω_d will it resonate?

28.19 ★★ *[Conceptual.]* A singer can shatter a wine glass by singing a sustained note at the right pitch but not at others, and only if the note is loud.

a. Explain in terms of natural frequency and resonance why only one pitch works.
b. Why must the note be loud, and what role does the glass's (small) damping play?

28.20 ★★★ *[Real-world application: engineering — the Mexico City problem.]* Reread the chapter opener. During the 1985 earthquake, the soft lake-bed clay beneath Mexico City shook with a dominant period of about $T_g = 2.0\ \text{s}$. Model a mid-rise building as a single mass–spring oscillator with an effective mass of $4.0 \times 10^6\ \text{kg}$ swaying on the effective stiffness k of its frame.

a. What effective stiffness k would give the building a natural period equal to the ground's, putting it at resonance?
b. A low building nearby has a natural period of about $0.3\ \text{s}$, and a tall tower one of about $5\ \text{s}$. Find the ratio of each natural frequency to the ground's frequency. Using Figure 28.4, explain why both fared better than the mid-rise building, even though all three were shaken by the same ground.
c. A modern tuned mass damper adds an auxiliary mass–spring oscillator tuned to the structure's natural frequency, but heavily damped, so that it absorbs energy near resonance. In one or two sentences, explain why adding damping at the resonant frequency is the key to keeping the amplitude finite. *(Refer back to Figure 28.4.)*

Challenge Problems

28.21 ★★★ *[Real-world application: seismology.]* A vertical-component seismometer is essentially a mass hanging on a spring: when the ground moves, the mass tends to stay put and the relative motion is recorded. One instrument uses a $0.50\ \text{kg}$ mass and is designed for a natural frequency of $f_0 = 1.0\ \text{Hz}$. Its damping force is $-b\ v$, and the amplitude of free oscillation decays as $A_0 e^{-bt/2m}$.

a. Find the spring constant k.
b. Critical damping occurs at $b_c = 2\sqrt{mk}$. Evaluate b_c.
c. The instrument is shipped with $b = 0.31\ \text{kg/s}$. Find the ratio b/b_c, the time $\tau = 2m/b$ for the amplitude to fall to $1/e$ of its initial value, the damped angular frequency $\omega' = \sqrt{\omega_0^2 - (b/2m)^2}$, and the number of complete oscillations that occur within one time τ.
d. Seismologists deliberately set the damping close to $b \approx 0.7\ b_c$ rather than leaving it as light as in part (c). Using your answer to (c), explain what goes wrong with a lightly damped seismometer when a sharp ground pulse arrives, and why near-critical damping gives a more faithful record.

28.22 ★★★ *[Real-world application: horology.]* A pendulum clock is regulated by a "seconds pendulum", a simple pendulum whose full period is $T = 2.000\ \text{s}$, so that it ticks once per second. The clock is calibrated in a laboratory where $g = 9.8000\ \text{m/s}^2$.

a. Find the required length ℓ.
b. The clock is moved to Denver, where $g = 9.7961\ \text{m/s}^2$. Find the new period. Then show that for a small change in g the fractional change in period is $\Delta T/T = -\tfrac{1}{2}\ \Delta g/g$, and use it to find how many seconds the clock gains or loses per day.
c. Back in the laboratory, the brass suspension rod ($\alpha = 19 \times 10^{-6}\ \text{K}^{-1}$) warms by $8.0\ \text{K}$ on a summer afternoon, so that $\Delta\ell/\ell = \alpha\ \Delta T_{\text{temp}}$. Find the resulting error in seconds per day, and state whether the clock now runs fast or slow.
d. A clockmaker corrects the summer error by raising the bob with an adjusting nut. By how many millimeters must the pendulum be shortened? Comment on why mechanical clock escapements are far less critical to accuracy than this one length.

28.23 ★★★ *[Real-world application: automotive.]* A car of total mass $1200\ \text{kg}$ rests on four identical springs, one at each wheel, and settles $3.0\ \text{cm}$ lower when its body is lowered onto them. Treat each corner as an independent mass–spring system carrying one quarter of the mass.

a. Find the spring constant k of one spring.
b. Find the natural frequency and period of the vertical bounce of one corner of the unloaded car.
c. Four passengers totaling $300\ \text{kg}$ get in, distributed evenly. Find the new natural frequency, and comment on the direction of the shift.
d. A stretch of "washboard" gravel road has ridges spaced $12\ \text{m}$ apart. At what speed (in m/s and km/hr) does the empty car's bounce hit resonance? At $25\ \text{m/s}$, is the car being driven above or below resonance, and what does your answer suggest about why drivers often report that a washboard road smooths out if they *speed up*?

This page titled 28.E: Oscillations (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29: WAVES I

Pluck a guitar string and it sings a definite note. Press it against a fret to shorten it and the note jumps higher; switch to a thicker, heavier string and the note drops lower; tighten the tuning peg and it rises again. A guitarist manages all of this by feel, but behind the music is a precise piece of physics. The string is not free to vibrate any old way: pinned at both ends, it can only oscillate in a discrete set of patterns, each with its own frequency, and the lowest of these is the note you hear. By the end of this chapter you will be able to predict that note from just three numbers: the length of the string, the tension in it, and how heavy it is. And you will see that the same analysis explains a violin, a piano, and the column of air in an organ pipe.

To get there we first have to understand what a wave *is*. When you pluck the string, you do not send any piece of the string traveling toward the listener; each bit of string just moves up and down in place. What travels is the *pattern* of that motion, and with it, energy. This is the essence of a wave: a disturbance that carries energy from one place to another without carrying matter along with it. The same idea describes ripples on a pond, the shock of an earthquake, the sound of a voice, and, as you saw in Chapter 27, light itself. We begin here with the most tangible case, a wave traveling along a stretched string, and build the mathematics that the rest of our study of waves will rely on.

Learning Objectives

- Distinguish transverse from longitudinal waves and pulse from periodic waves.
- Write the wave function of a sinusoidal traveling wave and relate the wave number, angular frequency, wavelength, period, and wave speed.
- Compute the speed of a wave on a string from its tension and linear mass density.
- Find the average power a sinusoidal wave transports along a string.
- Use phasors to add two waves of the same frequency and predict constructive and destructive interference.
- Describe standing waves and resonance, locate nodes and antinodes, and find the harmonic frequencies of a string fixed at both ends.

We will return to the guitar string, and to instruments generally, once we reach standing waves, and again in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

29.1: Wave Motion
29.2: The Wave Equation
29.3: Speed of a Wave
29.4: Energy of a Wave
29.5: Phasors
29.6: Standing Waves and Resonance
29.S: Waves I (Summary)
29.E: Waves I (Problems)

This page titled 29: Waves I is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.1: Wave Motion

A **mechanical wave** is a disturbance that travels through a material **medium**: a string, the air, the water, the ground. The defining feature is that the medium itself does not travel with the wave. As the wave passes, each particle of the medium is briefly displaced from equilibrium, passes the disturbance along to its neighbor, and then returns to where it started. Think of a stadium "wave": the pattern races around the arena, but every spectator stays in their own seat, merely standing and sitting in turn. What moves around the stadium is the *pattern*, and what the wave carries from place to place is *energy*, never matter.

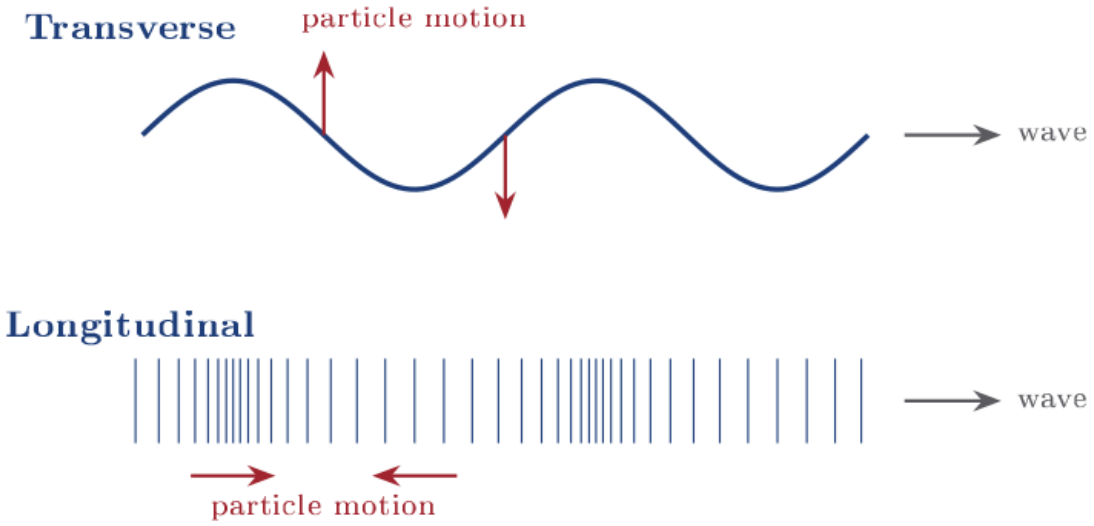


Figure 29.1. Two ways a wave can travel. In a *transverse* wave (top) the medium moves perpendicular to the wave's direction; in a *longitudinal* wave (bottom) the medium moves parallel to it, forming regions of compression and rarefaction. The red arrows show the motion at one instant: some particles are moving one way while others move the opposite way. In both, the wave moves to the right while each bit of the medium only oscillates back and forth about its own rest position.

•Figure description

Two wave types drawn one above the other, both traveling to the right. The upper panel, labeled Transverse, is a smooth sine curve along a string, with red arrows at two points showing the string moving straight up at one and straight down at the other -- perpendicular to the wave's direction. The lower panel, labeled Longitudinal, is a row of vertical lines whose spacing alternates between tightly bunched and widely spread, giving compressions and rarefactions, with red arrows below it showing the particle motion parallel to the wave's direction: at this instant the particles in a compression move to the right and the particles in a rarefaction move to the left.

It helps to sort waves along two independent lines. By their duration:

- A **pulse wave** is a single, short disturbance: one flick of a rope's end sends a lone hump traveling down it.
- A **periodic wave** results from a source repeating its motion over and over. When that source motion is simple harmonic (Chapter 28), the wave takes the shape of a sine curve, and we call it a **sinusoidal wave**. These are the waves we will analyze in detail.

And by the direction of the medium's motion relative to the wave's travel (Figure 29.1):

- In a **transverse wave** the particles of the medium move *perpendicular* to the direction the wave travels. A wave on a string is transverse: the string moves up and down while the wave moves sideways.
- In a **longitudinal wave** the particles move *parallel* to the direction of travel, bunching together and spreading apart. Sound, which we take up in Chapter 30, is the great example.

Checkpoint 29.1

A duck floats on a pond as ripples pass beneath it, traveling toward the shore. Does the duck drift toward the shore with the ripples, or does it mostly bob up and down in place? What does your answer tell you about whether the wave carries water or only energy across the pond?

Answer

Answers to Checkpoints

Checkpoint 29.1. It mostly bobs up and down (in fact it traces a small loop) in place; it does not ride the ripples to shore. Each bit of water moves back and forth about a fixed spot as the disturbance passes. The wave carries *energy* (and the pattern of the disturbance) across the pond, not the water itself.

This page titled 29.1: Wave Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.2: The Wave Equation

To describe a sinusoidal wave we need a function that gives the displacement y of the medium at every position x and every time t. A wave caused by simple harmonic motion and traveling in the $+x$ direction has the **wave function**

Key Equation: Sinusoidal traveling wave

$$y(x,t) = A\sin\left[\frac{2\pi}{\lambda}(x - vt) + \phi\right] = A\sin(kx - \omega t + \phi). \tag{29.1}$$

The two forms are identical once we introduce two abbreviations. The **wave number**

$$k \equiv \frac{2\pi}{\lambda} \tag{29.2}$$

counts radians of the wave pattern per meter, where the **wavelength** λ is the length of one full cycle of the wave in space. The **angular frequency**

$$\omega \equiv \frac{2\pi}{T} = 2\pi f \tag{29.3}$$

is exactly the quantity we met for oscillations: the source completes one cycle in the period T. The amplitude A is the maximum displacement, and the **phase constant** ϕ (just as in SHM) is fixed by the state of the medium at $x = 0$, $t = 0$. A wave moving in the $-x$ direction is written with a *plus* sign,

$$y(x,t) = A\sin(kx + \omega t + \phi) \qquad (\text{traveling in the } -x \text{ direction}). \tag{29.4}$$

Note: Which way is the wave going?

The sign is the opposite of what many students expect: *minus* ωt means the wave travels toward $+x$, and *plus* ωt means it travels toward $-x$. To see why, follow a single crest. A crest is a place where the phase $kx - \omega t$ has one particular value. As time goes on, ωt grows, so kx must grow by the same amount to keep the phase at that value: the crest moves to larger x. In $kx + \omega t$ the growing ωt must instead be offset by a shrinking kx, so the crest moves to smaller x.

The first form of equation (29.1) says the same thing. In $(x - vt)$ the displacement at position x and time t equals the displacement that was at $x - vt$ at time zero: the whole pattern has slid a distance vt toward $+x$. In $(x + vt)$ it has slid toward $-x$.

The wave speed follows from these definitions. In one period T the pattern advances exactly one wavelength λ, so the speed (the rate at which a crest moves) is

Key Equation: Wave speed

$$v = \frac{\lambda}{T} = f\lambda = \frac{\omega}{k}. \tag{29.5}$$

This relation holds for *every* periodic wave, of whatever kind. It is worth pausing on what it does and does not say. The speed v is the speed of the *pattern*, not of the particles of the medium; we will compute the particle speed separately in §29.4, and it is generally quite different.

Behind the sinusoidal solution lies a more general statement. Any traveling wave, regardless of its shape or type, satisfies the **wave equation**,

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\,\frac{\partial^2 y}{\partial t^2}, \tag{29.6}$$

a partial differential equation relating the curvature of the wave in space to its acceleration in time. The sinusoidal function (29.1) is one solution; you can confirm by differentiating that it satisfies equation (29.6) precisely when $v = \omega/k$, as it must. For a wave on a string y is the sideways displacement of the string; for a sound wave in a gas, as we will see, y stands instead for the back-and-forth displacement of the gas or the variation of its pressure.

Example 29.1: Reading a sinusoidal wave

A sinusoidal wave traveling in the $+x$ direction has amplitude $15.0\ \text{cm}$, wavelength $40.0\ \text{cm}$, and frequency $8.00\ \text{Hz}$. At $t = 0$ and $x = 0$ the displacement is $y = 15.0\ \text{cm}$, equal to the amplitude (the medium is at a crest there).

a. Find the wave number k, period T, angular frequency ω, and speed v.
b. Determine the phase constant ϕ and write the full wave function.

Solution

(a) From the definitions, taking care to convert to meters,

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{0.400\,\text{m}} = 15.7\,\text{rad/m}, \qquad T = \frac{1}{f} = \frac{1}{8.00\,\text{Hz}} = 0.125\,\text{s}, \tag{29.2.1}$$

$$\omega = 2\pi f = 2\pi(8.00\,\text{Hz}) = 50.3\,\text{rad/s}, \qquad v = f\lambda = (8.00\,\text{Hz})(0.400\,\text{m}) = 3.20\,\text{m/s}. \tag{29.2.2}$$

(b) At $x = 0$, $t = 0$ the wave function (29.1) reads $y = A\sin\phi$. Since $y = A$ there, we need $\sin\phi = 1$, so $\phi = \pi/2$. The full wave function is therefore

$$\boxed{y(x,t) = (0.150\,\text{m})\sin\left[(15.7\,\text{rad/m})x - (50.3\,\text{rad/s})t + \tfrac{\pi}{2}\right]} \tag{29.2.3}$$

which, because $\sin(\theta + \pi/2) = \cos\theta$, is the same as $y = (0.150\ \text{m})\cos(kx - \omega t)$: a cosine, peaking at the origin at $t = 0$, exactly as described.

This page titled 29.2: The Wave Equation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.3: Speed of a Wave

Equation (29.5), $v = f\lambda$, tells us how speed, frequency, and wavelength are linked, but it does not tell us what *sets* the speed in the first place. That is determined by the medium. For a wave on a string the speed depends on two properties of the string: how tightly it is stretched, and how heavy it is.

Key Equation: Speed of a wave on a string

$$v = \sqrt{\frac{F_T}{\mu}}, \tag{29.7}$$

where F_T is the tension in the string and $\mu = m/L$ is its **linear mass density**, the mass per unit length (units $\mathrm{kg/m}$). The form makes physical sense: a tighter string (larger F_T) snaps back faster and so carries the wave more quickly, while a heavier string (larger μ) has more inertia and carries it more slowly. Note what is *not* in the formula: the frequency. Tension and mass density alone fix the speed; the frequency at which you shake the string then determines the wavelength through $\lambda = v/f$.

Example 29.2: Signaling up a rescue cable

An $80.0\ \mathrm{kg}$ hiker is rescued by a helicopter that lowers a cable of mass $8.00\ \mathrm{kg}$ and length $15.0\ \mathrm{m}$. A sling of mass $70.0\ \mathrm{kg}$ hangs from the cable's end, and the hiker is clipped into the sling. The helicopter then accelerates straight up. To signal the pilot, the hiker sends transverse pulses up the cable; one pulse takes $0.250\ \mathrm{s}$ to travel the cable's length. Assuming the tension is uniform along the cable, what is the helicopter's upward acceleration?

Solution

First get the wave speed from the timing,

$$v = \frac{L}{t} = \frac{15.0\,\mathrm{m}}{0.250\,\mathrm{s}} = 60.0\,\mathrm{m/s}, \tag{29.3.1}$$

and the cable's linear density, $\mu = m_{\mathrm{cable}}/L = (8.00\ \mathrm{kg})/(15.0\ \mathrm{m}) = 0.533\ \mathrm{kg/m}$. Equation (29.7) then gives the tension:

$$F_T = \mu v^2 = (0.533\,\mathrm{kg/m})(60.0\,\mathrm{m/s})^2 = 1920\,\mathrm{N}. \tag{29.3.2}$$

This tension is the upward force the cable exerts on the load hanging below it: the sling plus the hiker, a mass $M = 70.0 + 80.0 = 150\ \mathrm{kg}$. (The "uniform tension" assumption lets us treat the cable's own weight as negligible in this force balance; its mass enters only through μ.) Newton's second law for the accelerating load, taking up as positive, gives $F_T - Mg = Ma$, so

$$a = \frac{F_T}{M} - g = \frac{1920\,\mathrm{N}}{150\,\mathrm{kg}} - 9.8\,\mathrm{m/s^2} = 12.8 - 9.8 \approx \boxed{3.0\,\mathrm{m/s^2}}. \tag{29.3.3}$$

The acceleration is positive, so the helicopter is indeed speeding upward, and the hiker learned it from nothing but a stopwatch: a pulse time turned into a tension, and the tension into an acceleration.

Checkpoint 29.2

(a) The tension in a string is quadrupled. By what factor does the wave speed change? (b) A second string has four times the linear density but the same tension. How does its wave speed compare? (c) You shake the end of the first string twice as fast. Does the wave speed change? What does change?

Answer

Answers to Checkpoints

Checkpoint 29.2. (a) The wave speed on a string is $v = \sqrt{F_T/\mu}$, so quadrupling F_T doubles v. (b) Quadrupling μ at the same tension halves v. (c) No. The speed is set by the string's tension and density alone, not by the source. Shaking twice as fast doubles the frequency, and since $v = f\lambda$ stays fixed, the wavelength halves.

This page titled 29.3: Speed of a Wave is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.4: Energy of a Wave

The entire purpose of a wave is to transport energy. For a wave on a string we can say exactly how much, and how fast. As the wave passes, each small element of string of mass dm oscillates up and down in SHM, and so carries kinetic energy $dK = \frac{1}{2}\ dm\ v_t^2$, where v_t is the element's *transverse* speed, not the wave speed v, but the speed at which the bit of string moves sideways. We find it by differentiating the wave function with respect to time at fixed position:

$$v_t = \frac{\partial y}{\partial t} = -\omega A \cos(kx - \omega t). \tag{29.8}$$

Adding up the kinetic and (equal, on average) potential energy carried past a point over one wavelength, and dividing by the period to get the rate of energy transport, gives the average power

$$P = \frac{E}{T} = \frac{1}{2}\mu\omega^2 A^2 \left(\frac{\lambda}{T}\right), \tag{29.4.1}$$

which, since $\lambda/T = v$, reduces to the compact result

Key Equation: Average power of a wave on a string

$$P = \tfrac{1}{2}\,\mu\,\omega^2 A^2 v. \tag{29.9}$$

The strong dependences are worth memorizing: the power a wave carries grows with the s*quare* of both the frequency and the amplitude. Doubling the amplitude of a wave on a string quadruples the energy it delivers; doubling the frequency does the same.

Example 29.3: Power to drive a string

A taut string with linear density $\mu = 5.00 \times 10^{-2}\ \text{kg/m}$ is under a tension of $80.0\ \text{N}$. How much power must be supplied to generate sinusoidal waves of frequency $60.0\ \text{Hz}$ and amplitude $6.00\ \text{cm}$?

Solution

We need v and ω before applying equation (29.9). The wave speed is

$$v = \sqrt{\frac{F_T}{\mu}} = \sqrt{\frac{80.0\,\text{N}}{5.00 \times 10^{-2}\,\text{kg/m}}} = \sqrt{1600} = 40.0\,\text{m/s}, \tag{29.4.2}$$

and the angular frequency is $\omega = 2\pi f = 2\pi(60.0\ \text{Hz}) = 377\ \text{rad/s}$. With $A = 0.0600\ \text{m}$,

$$P = \tfrac{1}{2}\mu\omega^2 A^2 v = \tfrac{1}{2}(5.00 \times 10^{-2})(377)^2(0.0600)^2(40.0) \approx \boxed{512\,\text{W}}. \tag{29.4.3}$$

About half a kilowatt: a reminder that shaking a string at $60\ \text{Hz}$ with a $6\ \text{cm}$ amplitude is hard work, because of those two squared factors in the power.

This page titled 29.4: Energy of a Wave is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.5: Phasors

What happens when two waves arrive at the same place at the same time? The answer is the **superposition principle**: the net displacement is simply the sum of the displacements each wave would produce alone. Where the two reinforce, we get **constructive interference**; where they oppose, **destructive interference**. To make this quantitative for sinusoidal waves, we introduce a wonderfully convenient bookkeeping device: the **phasor**.

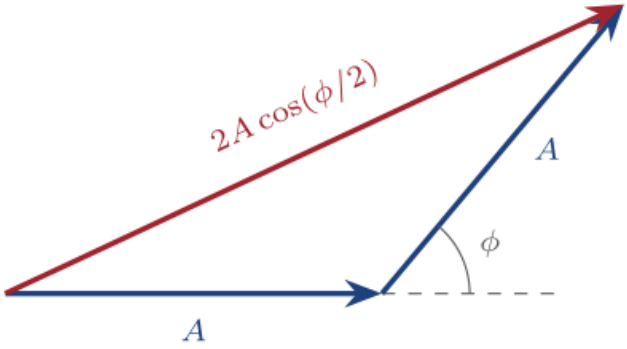


Figure 29.2. Adding two waves with a phasor diagram. Each wave is an arrow of length A; the second is turned through the phase difference ϕ and placed head to tail with the first. The resultant (red) has length $2A\cos(\phi/2)$: maximal when $\phi = 0$ (arrows aligned) and zero when $\phi = \pi$ (arrows opposed).

•Figure description

A phasor addition diagram. A first blue arrow of length A points to the right from the origin. A second blue arrow, also of length A, starts at the tip of the first and is rotated up by the phase difference phi. A red arrow from the origin to the tip of the second closes the triangle; it is the resultant, of length 2 A cosine of phi over 2. A small arc at the joint marks the angle phi.

A phasor represents a sinusoidal wave as an arrow of length equal to the wave's amplitude, oriented at an angle equal to the wave's phase. Adding two waves of the same frequency then becomes adding two arrows, head to tail, exactly like adding vectors (Figure 29.2). The length of the resultant arrow is the amplitude of the combined wave, and its angle is the combined phase. If phasors are new to you, or you met them in Chapter 26 and would like a refresher, Section C.8 of Appendix C builds them up step by step: how a rotating phasor traces out a sine curve, what a phase difference looks like, and why adding phasors adds sinusoids.

Consider two waves of equal amplitude A and frequency, traveling together but offset by a phase difference ϕ:

$$y_1 = A\sin(kx - \omega t), \qquad y_2 = A\sin(kx - \omega t + \phi). \tag{29.5.1}$$

By the superposition principle their sum is $y = y_1 + y_2 = A\big[\sin(kx - \omega t) + \sin(kx - \omega t + \phi)\big]$. Applying the trigonometric identity $\sin a + \sin b = 2\cos\left(\frac{a-b}{2}\right)\sin\left(\frac{a+b}{2}\right)$ (which is the algebraic statement of the phasor addition in Figure 29.2) gives

Key Equation: Superposition of two sinusoidal waves

$$y = \underbrace{2A\cos\left(\frac{\phi}{2}\right)}_{\text{amplitude}}\ \sin\left(kx - \omega t + \frac{\phi}{2}\right). \tag{29.10}$$

The result is still a sinusoidal wave of the same frequency and wavelength, but with a new amplitude $2A\cos(\phi/2)$ set entirely by the phase difference:

- If $\phi = 0$ (or $2\pi, 4\pi, \ldots$) the waves are *in phase*; $\cos(\phi/2) = 1$, the amplitude is $2A$, and the interference is fully **constructive**.
- If $\phi = \pi$ (or $3\pi, 5\pi, \ldots$) the waves are *out of phase*; $\cos(\phi/2) = 0$, the amplitude is zero, and the interference is fully **destructive**: the two waves cancel completely.

Checkpoint 29.3

Two identical waves are sent down the same string, the second lagging the first by a phase difference ϕ. For what value of ϕ (between 0 and 2π) is the combined wave's amplitude equal to A: the same as one wave alone? *(Set $2A\cos(\phi/2) = A$ and solve.)*

Answer

Answers to Checkpoints

Checkpoint 29.3. Setting $2A\cos(\phi/2) = A$ gives $\cos(\phi/2) = \frac{1}{2}$, so $\phi/2 = \pi/3$ and $\phi = 2\pi/3$ $(120°)$. The amplitude is really $2A\,|\cos(\phi/2)|$, since a negative value just flips the combined wave's sign, so $\cos(\phi/2) = -\frac{1}{2}$ works too, giving $\phi = 4\pi/3$ $(240°)$. In phasor terms, two equal phasors $120°$ apart form two sides of an equilateral triangle, so their sum has the same length as either one.

This page titled 29.5: Phasors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.6: Standing Waves and Resonance

We are finally ready to return to the guitar string. The key new ingredient is **reflection**: when a wave reaches the end of a string it bounces back. At a **fixed end** (the string tied down) the reflected wave is inverted; at a **free end** it is not. On a guitar string, pinned at both ends, a wave you launch reflects back and forth, and the original and reflected waves (identical but traveling in opposite directions) superpose.

Add a wave moving in $+x$ to an identical wave moving in $-x$:

$$y = A\sin(kx - \omega t) + A\sin(kx + \omega t) = 2A\sin(kx)\cos(\omega t), \tag{29.6.1}$$

again by the sum-to-product identity. The result is a **standing wave**:

Key Equation: Standing wave

$$y(x,t) = \left[2A\sin(kx)\right]\cos(\omega t). \tag{29.11}$$

This is not a traveling wave at all. Every point oscillates in time as $\cos(\omega t)$, but its *amplitude*, $2A\sin(kx)$, depends only on position. Some points never move: where $\sin(kx) = 0$, the amplitude is permanently zero. These are the **nodes**, located at

$$kx = n\pi \implies x = \frac{n\lambda}{2}, \qquad n = 0, 1, 2, \ldots \tag{29.6.2}$$

Halfway between them, where $\sin(kx) = \pm 1$, the amplitude reaches its maximum $2A$. These are the **antinodes**, at $x = \lambda/4, 3\lambda/4, 5\lambda/4, \ldots$. Nodes and antinodes alternate, each separated from the next of its kind by half a wavelength. Figure 29.3 shows three such patterns on a string tied at both ends, each at several instants during half a cycle. Which patterns a given string can hold, and at what frequencies, is the subject of the discussion of harmonics below.

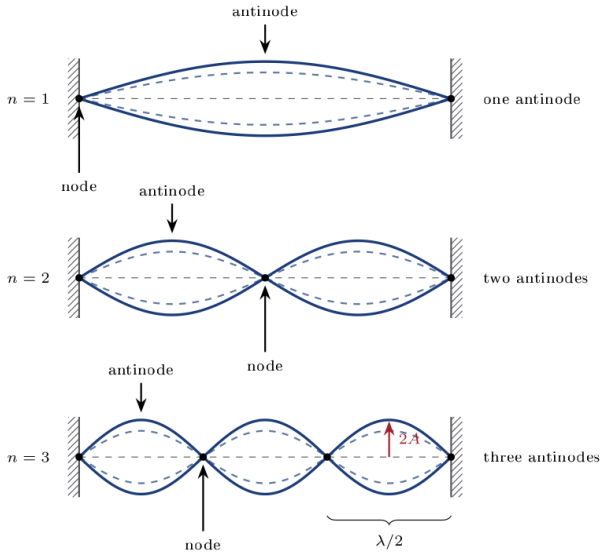


Figure 29.3. Three standing waves on a string tied at both ends, each drawn at five instants during half a cycle. The *nodes* (dots) never move; the *antinodes*, halfway between neighboring nodes, swing with the full amplitude $2A$. From top to bottom the string holds one, two, and three antinodes; in every case neighboring nodes are $\lambda/2$ apart. Every point between two neighboring nodes moves up and down together, and the next segment moves opposite to it. Nothing travels along the string.

•Figure description

Three standing-wave patterns on a string tied to walls at both ends, stacked one above another. Each pattern is drawn at five instants during half a cycle: a thick blue curve at its largest displacement, a dashed blue curve at about seventy percent of that, a dashed gray straight line where the string is momentarily flat, and the matching lighter and thick curves inverted. Black dots mark the nodes, where all five curves cross and the string never moves. Top, n equals 1: one antinode in the middle, with nodes only at the two ends. A labeled arrow points down to the antinode and another points up to the node at the left end. Middle, n equals 2: two antinodes and one node in the middle. Arrows point to the first antinode and to the middle node. Bottom, n equals 3: three antinodes and two interior nodes. Arrows point to the first antinode and to the first interior node. On the bottom pattern, a red arrow from the axis up to the thick curve at the last antinode is labeled 2 A, and a bracket below the axis between the second interior node and the right end is labeled lambda over 2.

Example 29.4: Nodes and antinodes of a standing wave

Two waves traveling in opposite directions on a string are $y_1 = (4.0)\sin(3.0x - 2.0t)$ and $y_2 = (4.0)\sin(3.0x + 2.0t)$, with x and y in centimeters and t in seconds.

a. Find the amplitude of the simple harmonic motion of the element at $x = 2.3$ cm.
b. Find the positions of the nodes and antinodes, taking one end of the string at $x = 0$.

Solution

The two waves superpose into the standing wave $y = [2A\sin(kx)]\cos(\omega t)$ with $A = 4.0$ cm, $k = 3.0$ rad/cm, and $\omega = 2.0$ rad/s.

(a) The amplitude at a point is $|2A\sin(kx)|$. At $x = 2.3$ cm,

$$2A\sin(kx) = 2(4.0)\sin\big[(3.0)(2.3)\big] = (8.0)\sin(6.9\text{ rad}) \approx (8.0)(0.578) = \boxed{4.6\,\text{cm}}. \tag{29.6.3}$$

(b) Nodes occur where $\sin(kx) = 0$, i.e. $kx = n\pi$:

$$x = \frac{n\pi}{3.0} = 0,\ 1.05,\ 2.09,\ 3.14,\ \ldots\ \text{cm}. \tag{29.6.4}$$

Antinodes occur where $\sin(kx) = \pm 1$, i.e. $kx = (n + \frac{1}{2})\pi$:

$$x = \frac{(n + \frac{1}{2})\pi}{3.0} = 0.52,\ 1.57,\ 2.62,\ \ldots\ \text{cm}, \tag{29.6.5}$$

each antinode sitting exactly halfway between two nodes, as expected.

Harmonics and resonance

A string fixed at both ends cannot vibrate at just any frequency. Both ends must be nodes, so a standing wave fits only if a whole number of half-wavelengths spans the length L: $L = n(\lambda/2)$, giving $\lambda_n = 2L/n$ (Figure 29.4a). Combined with $f = v/\lambda$, this picks out a discrete ladder of allowed frequencies, the **harmonics** or **resonant frequencies**:

Key Equation: Harmonics of a string fixed at both ends

$$f_n = n\,\frac{v}{2L}, \qquad n = 1, 2, 3, \ldots \tag{29.12}$$

The lowest, $f_1 = v/2L$, is the **fundamental** (the note you hear when you pluck the string) and the rest are its integer multiples. Driving the string at one of these frequencies produces a large-amplitude standing wave: this is **resonance**, the same phenomenon that brought down mid-rise buildings in the Mexico City earthquake of Chapter 28, here put to musical use. Now we can read off the guitarist's three controls: a shorter string (smaller L, from fretting), a tighter string (larger v, from tuning), or a lighter string (larger v, from a thinner gauge) all raise the pitch, exactly through equation (29.12) with $v = \sqrt{F_T/\mu}$.

A system that is *not* symmetric, such as a string or air column fixed at one end and free at the other, has a node at the fixed end and an antinode at the free end. It therefore fits only an odd number of quarter-wavelengths, $L = n(\lambda/4)$ with n odd, so $\lambda_n = 4L/n$ (Figure 29.4b), and its harmonics are

$$f_n = n\,\frac{v}{4L}, \qquad n = 1, 3, 5, \ldots, \tag{29.13}$$

containing only the *odd* multiples of the fundamental. Its fundamental, $v/4L$, is half that of a symmetric system of the same length, because only a quarter of a wavelength fits instead of a half. We will use this form for pipes in Chapter 30.

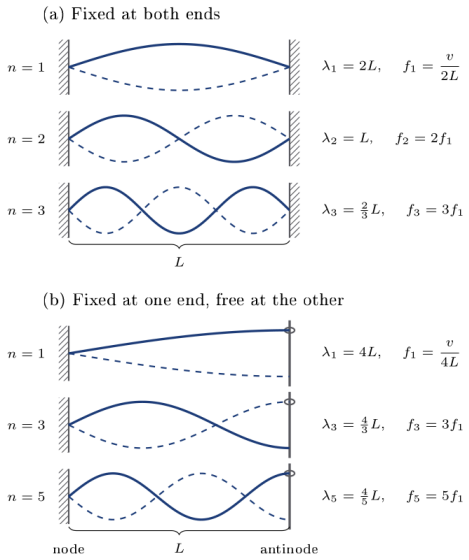


Figure 29.4. The lowest three allowed standing waves, drawn as the envelope of the string's motion (solid and dashed). (a) With both ends fixed, both ends are nodes and a whole number of half-wavelengths fits in L, so every integer multiple of $f_1 = v/2L$ is allowed. (b) With one end fixed and the other free (here a ring that slides on a frictionless rod), the fixed end is a node and the free end is an antinode, so an odd number of quarter-wavelengths fits in L. Only the odd multiples of $f_1 = v/4L$ are allowed.

•Figure description

Two panels of standing-wave patterns, each showing the lowest three allowed modes stacked vertically, drawn as the envelope of the string's motion: a solid blue curve and its dashed mirror image. Panel a, fixed at both ends: both ends are tied to walls and are nodes. The first mode is one loop, with lambda sub 1 equal to 2 L and f sub 1 equal to v over 2 L. The second mode is two loops, with lambda sub 2 equal to L and f sub 2 equal to 2 f sub 1. The third mode is three loops, with lambda sub 3 equal to 2 L over 3 and f sub 3 equal to 3 f sub 1. A bracket under the bottom string is labeled L. Panel b, fixed at one end and free at the other: the left end is tied to a wall and is a node; the right end is a ring free to slide up and down a vertical rod and is an antinode. The first mode is a quarter wavelength, half a loop, with lambda sub 1 equal to 4 L and f sub 1 equal to v over 4 L. The next allowed mode is one and a half loops, with lambda sub 3 equal to 4 L over 3 and f sub 3 equal to 3 f sub 1. The next is two and a half loops, with lambda sub 5 equal to 4 L over 5 and f sub 5 equal to 5 f sub 1. Even multiples are missing.

Example 29.5: Tuning a driven string

A string is tied to a sinusoidal oscillator at one end and runs over a support at the other, a distance $L = 1.20\ \text{m}$ away, where it is stretched by a hanging block of mass m. The string's linear density is $\mu = 1.6\ \text{g/m}$ and the oscillator's frequency is fixed at $120\ \text{Hz}$. Both ends are nodes.

a. What hanging mass m allows the oscillator to set up the *fourth* harmonic?
b. What standing-wave mode, if any, is set up if $m = 1.00\ \text{kg}$?

Solution

The block's weight is the tension, $F_T = mg$, so the wave speed is $v = \sqrt{mg/\mu}$, and the harmonic condition (29.12) reads

$$f_n = \frac{n}{2L}\sqrt{\frac{mg}{\mu}}. \tag{29.6.6}$$

(a) Setting $n = 4$, $f_4 = 120\ \text{Hz}$, and $\mu = 1.6 \times 10^{-3}\ \text{kg/m}$, solve for m:

$$\sqrt{\frac{mg}{\mu}} = \frac{2Lf_4}{n} = \frac{2(1.20)(120)}{4} = 72.0\,\text{m/s}, \tag{29.6.7}$$

so $mg/\mu = (72.0)^2 = 5184$, giving $F_T = mg = (5184)(1.6 \times 10^{-3}) = 8.29\ \text{N}$ and

$$m = \frac{F_T}{g} = \frac{8.29\,\text{N}}{9.8\,\text{m/s}^2} \approx \boxed{0.85\,\text{kg}}. \tag{29.6.8}$$

(b) With $m = 1.00\ \text{kg}$, the tension is $F_T = 9.8\ \text{N}$ and $v = \sqrt{F_T/\mu} = \sqrt{9.8/1.6 \times 10^{-3}} = 78.3\ \text{m/s}$. The mode number that the fixed $120\ \text{Hz}$ drive would require is

$$n = \frac{2Lf}{v} = \frac{2(1.20)(120)}{78.3} = 3.68, \tag{29.6.9}$$

which is *not* an integer. $\boxed{\text{No standing wave can form}}$ at this tension; only the discrete masses that make n a whole number produce resonance.

Checkpoint 29.4

A guitar string, fixed at both ends, has a fundamental frequency of $110\ \text{Hz}$. (a) What are the frequencies of the next two harmonics? (b) Where along the string are the nodes of the third harmonic? (c) A finger pressed lightly at the exact midpoint suppresses every mode that does not have a node there. Which frequency now sounds lowest?

Answer

Answers to Checkpoints

Checkpoint 29.4. (a) For a string fixed at both ends the harmonics are whole-number multiples of the fundamental, $f_n = nf_1$: the next two are $220\ \text{Hz}$ and $330\ \text{Hz}$. (b) The third harmonic has three half-wavelengths along the string, so besides the two fixed ends it has nodes one-third and two-thirds of the way along. (c) Only modes with a node at the midpoint survive: the even harmonics, $n = 2, 4, 6, \ldots$. The lowest is the second harmonic, $220\ \text{Hz}$, an octave above the open string. This is how guitarists play “harmonics.”

This page titled 29.6: Standing Waves and Resonance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.S: Waves I (Summary)

Summary

- A **mechanical wave** carries energy through a medium without carrying the medium with it. Waves are *transverse* (medium moves perpendicular to travel, e.g. a string) or *longitudinal* (parallel, e.g. sound), and *pulse* or *periodic*.
- A sinusoidal wave traveling toward $+x$ is $y(x,t) = A\sin(kx - \omega t + \phi)$; one traveling toward $-x$ has $+\omega t$ instead. The wave has **wave number** $k = 2\pi/\lambda$ and **angular frequency** $\omega = 2\pi/T = 2\pi f$. Its speed is $v = f\lambda = \omega/k$, the speed of the pattern (not of the medium). All traveling waves obey $\partial^2 y/\partial x^2 = \frac{1}{v^2}\partial^2 y/\partial t^2$.
- On a string, $v = \sqrt{F_T/\mu}$ with tension F_T and linear density $\mu = m/L$: independent of frequency.
- A wave on a string transports average power $P = \frac{1}{2}\mu\omega^2 A^2 v$, growing as the square of both frequency and amplitude.
- **Superposition**: displacements add. Two equal waves with phase difference ϕ combine (via **phasors**) to $y = 2A\cos(\phi/2)\sin(kx - \omega t + \phi/2)$: constructive at $\phi = 0$ (amplitude $2A$), *destructive* at $\phi = \pi$ (amplitude 0).
- Two opposite-traveling waves form a **standing wave** $y = [2A\sin(kx)]\cos(\omega t)$, with **nodes** at $x = n\lambda/2$ and **antinodes** between them.
- A string fixed at both ends resonates only at the **harmonics** $f_n = nv/2L$ ($n = 1, 2, 3, \ldots$); a system fixed at one end only, at $f_n = nv/4L$ (n odd).

Key Equations

Sinusoidal traveling wave

$$\begin{aligned} y(x,t) &= A\sin(kx - \omega t + \phi) \qquad (\text{toward } +x) \\ y(x,t) &= A\sin(kx + \omega t + \phi) \qquad (\text{toward } -x) \end{aligned} \tag{29.S.1}$$

Wave number and angular frequency

$$k = \frac{2\pi}{\lambda}, \qquad \omega = \frac{2\pi}{T} = 2\pi f \tag{29.S.2}$$

Wave speed

$$v = \frac{\lambda}{T} = f\lambda = \frac{\omega}{k} \tag{29.S.3}$$

The wave equation

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} \tag{29.S.4}$$

Speed of a wave on a string

$$v = \sqrt{\frac{F_T}{\mu}}, \qquad \mu = \frac{m}{L} \tag{29.S.5}$$

Transverse speed of a string element

$$v_t = \frac{\partial y}{\partial t} = -\omega A\cos(kx - \omega t) \tag{29.S.6}$$

Average power of a wave on a string

$$P = \tfrac{1}{2}\,\mu\,\omega^2 A^2 v \tag{29.S.7}$$

Superposition of two sinusoidal waves

$$y = 2A\cos\left(\frac{\phi}{2}\right)\sin\left(kx - \omega t + \frac{\phi}{2}\right) \tag{29.S.8}$$

Standing wave

$$y(x,t) = \left[\,2A\sin(kx)\,\right]\cos(\omega t) \tag{29.S.9}$$

Nodes and antinodes

$$\begin{aligned} x_{\text{node}} &= \frac{n\lambda}{2}, & n &= 0, 1, 2, \ldots \\ x_{\text{antinode}} &= \frac{(2n+1)\lambda}{4}, & n &= 0, 1, 2, \ldots \end{aligned} \tag{29.S.10}$$

Harmonics, both ends fixed

$$f_n = n\,\frac{v}{2L}, \qquad n = 1, 2, 3, \ldots \tag{29.S.11}$$

Harmonics, one end fixed

$$f_n = n\,\frac{v}{4L}, \qquad n = 1, 3, 5, \ldots \tag{29.S.12}$$

This page titled 29.S: Waves I (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

29.E: Waves I (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Take $g = 9.8\ \text{m/s}^2$ unless told otherwise.

Section 29.1 — Wave Motion

29.1 ★ *[Conceptual.]* Classify each wave as transverse or longitudinal, and as pulse or periodic: (a) a single flick sent down a stretched rope; (b) a sound wave from a steadily ringing bell; (c) the ripples spreading steadily from a dripping faucet; (d) the side-to-side wave on a long, continuously shaken spring.

29.2 ★ *[Conceptual.]* A cork floats on water as a steady train of waves passes. Describe the cork's motion, and explain what the wave transports past the cork and what it does not.

Section 29.2 — The Wave Equation

29.3 ★ A wave has wavelength $0.50\ \text{m}$ and frequency $120\ \text{Hz}$.

a. Find the wave number k.
b. Find the angular frequency ω.
c. Find the wave speed.

29.4 ★★ A transverse wave on a string is described by

$$y(x,t) = (0.020\,\text{m})\sin\!\left[(6.0\,\text{rad/m})x - (300\,\text{rad/s})t\right]. \tag{29.E.1}$$

a. State the amplitude, wave number, and angular frequency.
b. Find the wavelength, frequency, period, and speed.
c. In which direction does the wave travel?

29.5 ★★ A sinusoidal wave traveling in the $+x$ direction has amplitude $8.0\ \text{cm}$, wavelength $80.0\ \text{cm}$, and frequency $3.0\ \text{Hz}$. At $t = 0$ and $x = 0$ the displacement is zero and increasing. Write the full wave function, including the phase constant.

Section 29.3 — Speed of a Wave

29.6 ★ A string of linear density $4.0 \times 10^{-3}\ \text{kg/m}$ is under a tension of $36\ \text{N}$. Find the speed of transverse waves on it.

29.7 ★★ A $2.0\ \text{m}$ rope of mass $0.12\ \text{kg}$ is stretched with a tension of $60\ \text{N}$.

a. Find the wave speed.
b. If the rope is shaken at $8.0\ \text{Hz}$, what is the wavelength of the resulting wave?

29.8 ★★ *[Real-world application: engineering.]* A steel guy-wire of linear density $0.025\ \text{kg/m}$ carries a transverse pulse that travels its $12\ \text{m}$ length in $0.18\ \text{s}$. Find (a) the wave speed and (b) the tension in the wire.

29.9 ★★ A string has linear mass density $5.0\ \text{g/m}$ and is held under $80\ \text{N}$ of tension.

a. Find the speed of transverse waves on it.
b. What tension would be needed to double that speed?

Section 29.4 — Energy of a Wave

29.10 ★★ A string of linear density $0.040\ \text{kg/m}$ carries a wave of amplitude $5.0\ \text{cm}$ and angular frequency $200\ \text{rad/s}$ traveling at $30\ \text{m/s}$. Find the average power transported.

29.11 ★★ A wave on a string transports $120\ \mathrm{W}$. Without changing the string, tension, or frequency, you double the amplitude. What power does it now transport, and why?

29.12 ★★★ *[Real-world application: engineering.]* A string with $\mu = 6.0 \times 10^{-2}\ \mathrm{kg/m}$ under tension $150\ \mathrm{N}$ is to carry $300\ \mathrm{W}$ of power at a frequency of $50\ \mathrm{Hz}$. Find (a) the wave speed and (b) the required amplitude.

29.13 ★★ A sinusoidal wave travels on a string ($\mu = 20\ \mathrm{g/m}$) at $30\ \mathrm{m/s}$ with amplitude $8.0\ \mathrm{mm}$ and frequency $50\ \mathrm{Hz}$. Find the average power it transmits.

Section 29.5 — Phasors

29.14 ★ Two identical waves of amplitude $3.0\ \mathrm{cm}$ travel together on a string with a phase difference of ϕ. Find the amplitude of the resultant wave when (a) $\phi = 0$, (b) $\phi = \pi/2$, (c) $\phi = \pi$.

29.15 ★★ Two waves, each of amplitude A, interfere. The resultant amplitude is measured to be $1.4A$. Find the phase difference ϕ between them.

29.16 ★★ *[Conceptual.]* Using the phasor picture of Figure 29.2, explain why the resultant amplitude of two equal waves can never exceed $2A$, and identify the phase difference that achieves this maximum.

Section 29.6 — Standing Waves and Resonance & the Guitar String

29.17 ★ A string fixed at both ends is $0.80\ \mathrm{m}$ long, and transverse waves travel on it at $200\ \mathrm{m/s}$.

a. Find the fundamental frequency.
b. Find the frequencies of the second and third harmonics.

29.18 ★★ Two waves traveling in opposite directions, $y_1 = (2.0)\sin(4.0x - 3.0t)$ and $y_2 = (2.0)\sin(4.0x + 3.0t)$ (SI-scaled, x in m), form a standing wave.

a. Write the standing-wave function.
b. Find the positions of the first three nodes (for $x \geq 0$).
c. Find the amplitude of the element at $x = 0.30\ \mathrm{m}$.

29.19 ★★ *[Real-world application: music — the guitar string.]* A guitar string $0.65\ \mathrm{m}$ long has linear density $5.0 \times 10^{-3}\ \mathrm{kg/m}$ and is tuned by adjusting its tension.

a. What tension gives a fundamental frequency of $330\ \mathrm{Hz}$ (the note E)?
b. A guitarist frets the string, shortening its vibrating length to $0.49\ \mathrm{m}$ at the same tension. What is the new fundamental frequency?
c. To raise the open-string pitch by switching to a thinner string at the same length and tension, should the new string's linear density be larger or smaller? Explain using $f_1 = \frac{1}{2L}\sqrt{F_T/\mu}$.

29.20 ★★★ *[Driven string.]* A string of length $1.5\ \mathrm{m}$ and linear density $2.0 \times 10^{-3}\ \mathrm{kg/m}$, fixed at both ends, is driven by an oscillator at $140\ \mathrm{Hz}$ while a hanging block sets the tension.

a. What tension is needed to set up the third harmonic?
b. What hanging mass does this require?
c. If the tension is doubled, can a standing wave still be set up at $140\ \mathrm{Hz}$? If so, which harmonic; if not, explain.

Challenge Problems

29.21 ★★★ *[Real-world application: engineering — structural health monitoring.]* The tension in a cable-stayed bridge's stay cables cannot be measured directly once the bridge is built, so engineers pluck a cable, record its vibration with an accelerometer, and infer the tension from the pitch. One stay is $48\ \mathrm{m}$ long between its anchorages, has a mass of $1150\ \mathrm{kg}$, and is effectively fixed at both ends. Its measured fundamental frequency is $2.15\ \mathrm{Hz}$.

a. Find the linear mass density μ and the wave speed v on the cable.
b. Find the tension in the cable, in newtons and in meganewtons.
c. Show that $F_T \propto f_1^2$, so that a fractional loss of tension shows up as a drop in pitch. If corrosion of an anchorage lets the tension fall by 10%, what fundamental frequency would the accelerometer record?
d. The accelerometer records usable data up to $10\ \text{Hz}$. How many harmonics of the original cable fall in that band, and what are their frequencies?

29.22 ★★★ *[Real-world application: music — natural harmonics.]* A guitarist plays a "natural harmonic" by resting a finger lightly on a vibrating string (just touching it, not pressing it to the fret), which forces a node at the point of contact without changing the string's length. One string is $0.65\ \text{m}$ long between nut and bridge and has a fundamental frequency of $196\ \text{Hz}$.

a. Find the speed of transverse waves on the string.
b. The guitarist touches the string one third of the way along, at $x = \ell/3$. Which harmonics n have a node there? What is the lowest surviving frequency, and which musical interval above the open string is it?
c. Repeat for a light touch at the midpoint, $x = \ell/2$, and for a touch at $x = \ell/4$.
d. A harmonic played this way sounds thinner and purer than a fretted note of the same pitch. Explain this in terms of which harmonics the light touch suppresses, and why pressing the string to the fret at $x = \ell/3$ instead produces a different pitch.

29.23 ★★★ *[Real-world application: engineering — a hanging mooring line.]* A uniform rope of length L and mass M hangs vertically from a ship's davit with nothing attached to its lower end. A sailor flicks the bottom of the rope, sending a transverse pulse up it. Because each element of the rope must support only the weight of the rope *below* it, the tension is not constant.

a. Let y be the height above the free lower end. Show that the tension there is $F_T(y) = \mu g y$, where $\mu = M/L$, and hence that the pulse speed is $v(y) = \sqrt{gy}$: independent of the rope's mass.
b. Using $v = dy/dt$, set up and evaluate the integral for the time the pulse takes to travel the full length, and show that $t = 2\sqrt{L/g}$.
c. Evaluate t for a $12\ \text{m}$ rope.
d. Compare your answer with the (incorrect) estimate you would get by assuming a uniform tension equal to the rope's full weight Mg throughout. Express the ratio of the two answers exactly, and explain physically which part of the journey the uniform-tension estimate gets most wrong.

This page titled 29.E: Waves I (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30: WAVES II

You are standing at the curb when an ambulance races past, siren blaring. As it approaches you the pitch is high and urgent; the instant it passes, the pitch drops noticeably, and it keeps falling as the ambulance recedes. The siren itself never changed its note: a paramedic riding inside would hear one steady pitch the whole time. What changed is the relationship between you and the source of the sound, and that relationship rewrote the frequency that reached your ear. By the end of this chapter you will be able to predict exactly how much the pitch drops, and to turn the effect around: from the size of the pitch change, you could clock the ambulance's speed. The same physics lets a radar gun catch a speeder, a doctor image a beating heart, and an astronomer measure the expansion of the universe.

This is a chapter about sound: the most familiar wave in daily life, and the prototype of a longitudinal wave. We will find how fast sound travels and why, what happens when two sound waves overlap, how we measure loudness on the peculiar logarithmic scale of decibels, and why two slightly mistuned notes produce the throbbing we call beats. We close with the Doppler effect of the ambulance siren. Throughout, the wave ideas of Chapter 29 (wave speed, superposition, interference) carry over directly; only the medium has changed.

Learning Objectives

- Describe a sound wave as a longitudinal wave of pressure and displacement.
- Compute the speed of sound from the medium's bulk modulus and density.
- Predict constructive and destructive interference of sound from the path-length difference between two sources.
- Relate the intensity of a sound to its distance from the source, and convert intensity to the decibel scale.
- Find the beat frequency produced by two nearly equal tones.
- Apply the Doppler effect to moving sources and observers, choosing the signs correctly.

We will come back to the ambulance siren when we reach the Doppler effect, and in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

30.1: Sound Waves
30.2: Speed of Sound
30.3: Wave Interference
30.4: Intensity and Decibels
30.5: Beats
30.6: The Doppler Effect
30.S: Waves II (Summary)
30.E: Waves II (Problems)

This page titled 30: Waves II is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.1: Sound Waves

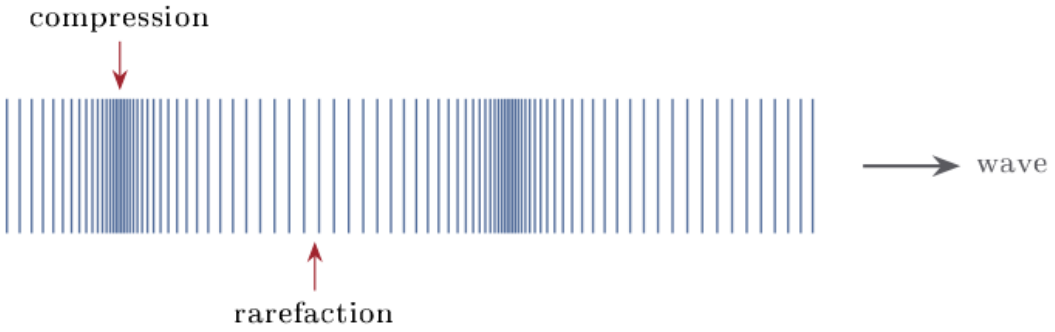


Figure 30.1. A sound wave in a gas. The medium is alternately squeezed into *compressions* (high pressure) and stretched into *rarefactions* (low pressure), and this pattern travels to the right while each particle merely oscillates back and forth along the direction of travel: a longitudinal wave.

•Figure description

A sound wave in a gas, drawn as a long row of closely spaced vertical lines representing the gas particles. The spacing varies smoothly along the row: in some places the lines crowd together, marked compression, and in between they spread apart, marked rarefaction. An arrow at the right shows the pattern traveling to the right, while each particle only oscillates back and forth along that same direction.

A **sound wave** is the most common longitudinal wave. As it travels through a medium (air, water, steel) it pushes the particles of the medium back and forth along the direction of travel, creating a moving pattern of **compressions** (regions of high pressure, where particles bunch together) and **rarefactions** (regions of low pressure, where they spread apart), as in Figure 30.1. Your eardrum responds to the arriving pressure swings, and your brain interprets them as sound.

Because a sound wave is at once a displacement of the medium and a variation of its pressure, we can describe it either way. The displacement of a small element of the medium from its equilibrium position is

$$s(x,t) = A\cos(kx - \omega t), \tag{30.1}$$

while the **pressure variation** about the equilibrium pressure is

$$\Delta P = \Delta P_{\max}\sin(kx - \omega t), \qquad \Delta P_{\max} = \rho\, v\, \omega A. \tag{30.2}$$

The two descriptions are a quarter-cycle out of step: where the displacement is greatest, the pressure variation is momentarily zero, and vice versa. The pressure amplitude $\Delta P_{\max}$ grows with the density ρ of the medium, the wave speed v, the frequency (through ω), and the displacement amplitude A.

Sound waves are everywhere useful precisely because they carry information and energy through opaque media that light cannot penetrate. We use them for communication and music, for echolocation by bats and ships' sonar, for finding oil and mapping the sea floor, and for medical imaging: the ultrasound that pictures a fetus is a sound wave far above the range of human hearing.

This page titled 30.1: Sound Waves is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.2: Speed of Sound

Sound travels through solids, liquids, and gases, and in each its speed is set by two properties of the medium: how stiff it is (how strongly it resists compression) and how dense it is. The speed is

Key Equation: Speed of sound

$$v = \sqrt{\frac{B}{\rho}}, \tag{30.3}$$

where B is the **bulk modulus** of the medium (the pressure change required to produce a given fractional change in volume, a measure of stiffness) and ρ is its density. Notice the close kinship with the speed of a wave on a string, $v = \sqrt{F_T/\mu}$ from Chapter 29: in both, the speed is the square root of a "stiffness" (restoring strength) divided by an "inertia" (mass). A stiffer medium carries sound faster; a denser one, slower. In air at room temperature the speed of sound is about $343\ \mathrm{m/s}$, a value we will use throughout the chapter; in water it is roughly $1500\ \mathrm{m/s}$, and in steel faster still, because solids are so much stiffer than gases.

Checkpoint 30.1

Sound travels several times faster in water than in air, even though water is far denser. Since a larger density ρ in equation (30.3) would by itself *lower* the speed, what must be true of water's bulk modulus B compared with air's, and why does that make physical sense?

Answer

Answers to Checkpoints

Checkpoint 30.1. The speed is $v = \sqrt{B/\rho}$. Water is about 1000 times denser than air, which by itself would make sound in water about 30 times slower. Since sound in water is actually several times *faster*, water's bulk modulus must be enormously larger than air's, in fact about $2 \times 10^9\ \mathrm{Pa}$ versus about $1.4 \times 10^5\ \mathrm{Pa}$, some $15\ 000$ times larger. That makes sense: a gas is easily squeezed, while water is nearly incompressible, so a compression in water pushes back far more strongly and passes along far faster.

This page titled 30.2: Speed of Sound is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.3: Wave Interference

When two sound waves overlap, they obey the superposition principle of Chapter 29: the pressure variations add. Whether the result is loud or quiet at a given point depends on the **phase difference** between the two waves when they arrive, and for two sources emitting in step, that phase difference is set by the difference in the distances the waves have traveled, the **path-length difference**

$$\Delta r = |r_2 - r_1|, \tag{30.3.1}$$

where r_1 and r_2 are the distances from the two sources to the listener (Figure 30.2a). The two waves leave their sources in step, but the wave with the longer path has to cover the extra distance Δr, so it arrives having gone through extra cycles. What matters is how many wavelengths fit into that extra distance (Figure 30.2b). This idea returns for light in Chapters 33 and 34, where it decides which places on a screen are bright and which are dark.

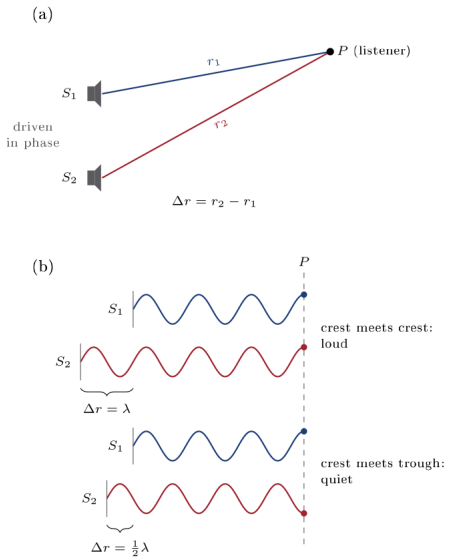


Figure 30.2. Path-length difference. (a) Two speakers driven in phase send the same tone to a listener at P, along paths of lengths r_1 and r_2. (b) The same two paths straightened out and lined up so that both end at P. The waves leave their speakers in step, so what they are doing when they reach P depends only on the extra distance Δr that the second wave covers. When Δr is a whole wavelength (top), the waves arrive in step, crest with crest, and reinforce. When Δr is half a wavelength (bottom), a crest arrives with a trough, and they cancel.

•Figure description

Two panels about path-length difference. Panel a: two loudspeakers, S sub 1 and S sub 2, one above the other at the left, and a listener P up and to the right. A straight line from S sub 1 to P is labeled r sub 1; a longer line from S sub 2 to P is labeled r sub 2. Text below reads delta r equals r sub 2 minus r sub 1. Panel b: the two paths straightened out and lined up so that both end at P, marked by a vertical dashed line at the right. Each path carries a sine wave that starts at its speaker. In the top pair the lower path is longer by one wavelength, marked with a bracket labeled delta r equals lambda, and at P both waves are at a crest: crest meets crest, loud. In the bottom pair the lower path is longer by half a wavelength, bracket labeled delta r equals lambda over 2, and at P the upper wave is at a crest while the lower wave is at a trough: crest meets trough, quiet.

A path-length difference of one whole wavelength puts the waves back in step, so the conditions are:

Key Equation: Interference from a path difference

$$\text{constructive (loud):} \quad \Delta r = m\lambda, \tag{30.4}$$

$$\text{destructive (quiet):} \quad \Delta r = \left(m + \tfrac{1}{2}\right)\lambda, \tag{30.5}$$

with $m = 0, 1, 2, \ldots$. If the two paths differ by a whole number of wavelengths the crests arrive together and reinforce; if they differ by a half-integer number, a crest from one meets a trough from the other and they cancel.

Example 30.1: Interference in a tunable sound path

Sound from a speaker divides into two equal parts that travel separate paths and recombine at a receiver; one path length can be varied by sliding a section, like a trombone slide. The speed of sound is $343\ \mathrm{m/s}$ and the speaker frequency is $536\ \mathrm{Hz}$.

a. Starting from a position of constructive interference, by what minimum distance must the slider be moved to produce destructive interference instead?
b. What minimum slider displacement (from the original position) restores constructive interference?

Solution

The wavelength of the sound is

$$\lambda = \frac{v}{f} = \frac{343\,\mathrm{m/s}}{536\,\mathrm{Hz}} = 0.640\,\mathrm{m}. \tag{30.3.2}$$

Moving the slider by a distance d lengthens that branch of the path on both the outgoing and returning legs, so the path-length difference changes by $\Delta r = 2d$.

(a) To go from constructive to destructive interference the path difference must change by half a wavelength, $2d = \lambda/2$, so

$$d = \frac{\lambda}{4} = \frac{0.640\,\mathrm{m}}{4} \approx \boxed{0.160\,\mathrm{m}}. \tag{30.3.3}$$

(b) To return to constructive interference the path difference must change by a full wavelength, $2d = \lambda$, so

$$d = \frac{\lambda}{2} = \frac{0.640\,\mathrm{m}}{2} \approx \boxed{0.320\,\mathrm{m}}. \tag{30.3.4}$$

The factor of two (because the slider adds to the path twice) is the easy step to overlook here.

Checkpoint 30.2

Two loudspeakers driven in phase emit the same tone, of wavelength $1.0\ \mathrm{m}$. (a) You stand $3.0\ \mathrm{m}$ from one speaker and $4.5\ \mathrm{m}$ from the other. Is the sound loud or quiet there? (b) You move to a spot $3.0\ \mathrm{m}$ from one and $5.0\ \mathrm{m}$ from the other. Loud or quiet?

Answer

Answers to Checkpoints

Checkpoint 30.2. (a) The path difference is $4.5 - 3.0 = 1.5\ \mathrm{m} = 1.5\lambda$, a whole number plus one half, so the waves arrive half a cycle out of step: destructive interference, a *quiet* spot. (b) Now the path difference is $2.0\ \mathrm{m} = 2\lambda$, a whole number of wavelengths, so the waves arrive in step: constructive interference, a *loud* spot.

This page titled 30.3: Wave Interference is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.4: Intensity and Decibels

The loudness of a sound is governed by its **intensity** I, the power the wave delivers per unit area of surface facing it:

$$I = \frac{P}{A}. \tag{30.6}$$

A source that radiates power P uniformly in all directions spreads it over the surface of an expanding sphere of area $4\pi r^2$, so at distance r the intensity is

Key Equation: Intensity from a point source

$$I = \frac{P}{4\pi r^2}. \tag{30.7}$$

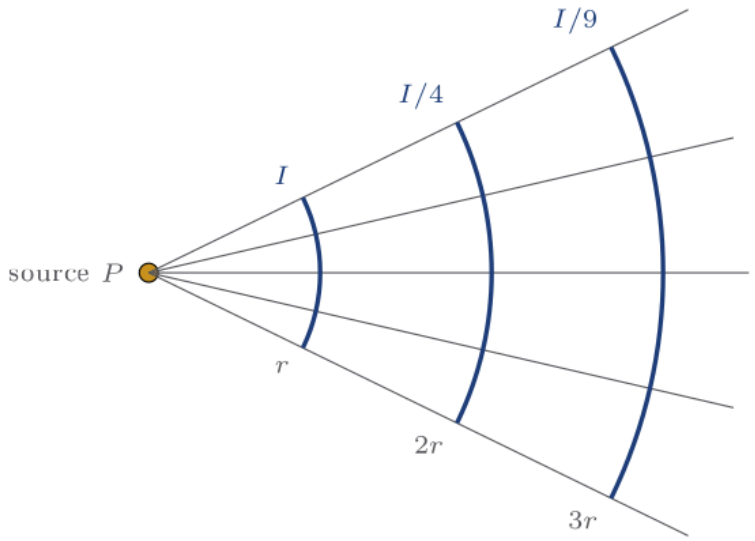


Figure 30.3. The inverse-square law. A point source radiates its power P over an expanding sphere of area $4\pi r^2$, so the intensity falls as $1/r^2$: at twice the distance the intensity is one-fourth, at three times the distance one-ninth.

•Figure description

A point source at the left radiating power P, with straight rays fanning out to the right. Three arcs cross the rays at increasing distances r, 2 r and 3 r. The nearest arc is labeled intensity I, the middle one I over 4, and the farthest I over 9 -- the same power spread over an area that grows as the square of the distance.

This is an *inverse-square law*: double your distance from the source and the intensity drops to a quarter, because the same power is spread over four times the area (Figure 30.3). The relation also lets us compare two distances directly, $I_2/I_1 = (r_1/r_2)^2$, which is often the quickest route in a problem.

The human ear, however, does not perceive loudness in proportion to intensity. It responds roughly *logarithmically*: each multiplication of the intensity by ten is heard as the same step up in loudness. To match this, we measure sound on the **decibel** scale. The **sound intensity level** β, in decibels (dB), is

Key Equation: Sound intensity level

$$\beta = (10\ \text{dB}) \log\left(\frac{I}{I_0}\right), \qquad I_0 = 10^{-12}\,\text{W/m}^2, \tag{30.8}$$

where the logarithm is base ten (not the natural log) and I_0 is the **reference intensity**, the faintest sound a healthy ear can detect, taken as 0 dB. The scale's logarithmic nature has a tidy consequence: multiplying the intensity by 10 adds exactly 10 dB, and

multiplying it by 2 adds about 3 dB. So a $60\ \mathrm{dB}$ conversation is not “twice” a $30\ \mathrm{dB}$ whisper but a thousand times more intense.

Example 30.2: Intensity of a point source

A point source emits sound waves with an average power output of $80.0\ \mathrm{W}$.

a. Find the intensity $3.00\ \mathrm{m}$ from the source.
b. At what distance is the intensity $1.00 \times 10^{-8}\ \mathrm{W/m^2}$?

Solution

(a) Using equation (30.7) with $r = 3.00\ \mathrm{m}$,

$$I = \frac{P}{4\pi r^2} = \frac{80.0\,\mathrm{W}}{4\pi(3.00\,\mathrm{m})^2} \approx \boxed{0.707\,\mathrm{W/m^2}}. \tag{30.4.1}$$

(b) Solving the same relation for r,

$$r = \sqrt{\frac{P}{4\pi I}} = \sqrt{\frac{80.0\,\mathrm{W}}{4\pi(1.00 \times 10^{-8}\,\mathrm{W/m^2})}} \approx 2.52 \times 10^{4}\,\mathrm{m} = \boxed{25.2\,\mathrm{km}}. \tag{30.4.2}$$

Sound spreads thin remarkably fast: the intensity has fallen by eight orders of magnitude over $25\ \mathrm{km}$, just as the inverse-square law demands.

Example 30.3: Two machines and the decibel scale

The intensity of sound delivered by a single operating machine at a worker’s location is $2.0 \times 10^{-7}\ \mathrm{W/m^2}$.

a. Find the sound level in decibels when one machine runs.
b. Find the sound level when a second, identical machine runs alongside it.

Solution

(a) With $I = 2.0 \times 10^{-7}\ \mathrm{W/m^2}$ and $I_0 = 10^{-12}\ \mathrm{W/m^2}$,

$$\beta_1 = (10)\log\left(\frac{2.0 \times 10^{-7}}{1.0 \times 10^{-12}}\right) = (10)\log(2.0 \times 10^{5}) \approx \boxed{53\,\mathrm{dB}}. \tag{30.4.3}$$

(b) Two machines deliver twice the intensity, $I = 4.0 \times 10^{-7}\ \mathrm{W/m^2}$, so

$$\beta_2 = (10)\log\left(\frac{4.0 \times 10^{-7}}{1.0 \times 10^{-12}}\right) = (10)\log(4.0 \times 10^{5}) \approx \boxed{56\,\mathrm{dB}}. \tag{30.4.4}$$

Doubling the intensity raised the level by only about 3 dB, not by a factor of two: the logarithm at work, and the reason a second machine sounds only slightly louder than one.

Checkpoint 30.3

(a) A sound’s intensity increases by a factor of 100. By how many decibels does its level rise? (b) How many times more intense is a $90\ \mathrm{dB}$ concert than a $60\ \mathrm{dB}$ conversation? (c) You move three times farther from a small source. By about how many decibels does the level drop?

Answer

Answers to Checkpoints

Checkpoint 30.3. (a) The level changes by $10\log_{10}(I_2/I_1) = 10\log_{10}100 = 10(2) = 20\ \mathrm{dB}$. (b) A difference of $30\ \mathrm{dB}$ means

$10\log_{10}(I_2/I_1) = 30$, so $I_2/I_1 = 10^3 = 1000$: the concert is a thousand times more intense. (c) For a point source $I \propto 1/r^2$, so tripling the distance divides the intensity by 9. The level drops by $10\log_{10} 9 \approx 9.5\ \text{dB}$.

This page titled 30.4: Intensity and Decibels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.5: Beats

So far we have combined waves of the *same* frequency. Sound two notes of *slightly different* frequency together and something new happens: the combined loudness rises and falls in a slow throb. At one moment the two waves are in phase and reinforce (loud); a fraction of a second later they have drifted out of phase and cancel (soft); then back again. This periodic swelling is called **beats**, and the rate at which the loudness pulses is the **beat frequency**.

Key Equation: Beat frequency

$$f_{\text{beat}} = |f_2 - f_1|. \tag{30.9}$$

The beat frequency is simply the difference of the two frequencies. This is exactly how a musician tunes by ear: playing a note against a reference, they adjust until the beats slow and finally vanish, at which point the two frequencies are equal.

Example 30.4: Finding a tuning fork's frequency by beats

Two out-of-tune flutes play the same intended note. One sounds at $246\ \text{Hz}$ and the other at $281\ \text{Hz}$. A tuning fork sounded with the $246\ \text{Hz}$ flute produces a beat frequency of $15.0\ \text{Hz}$; sounded with the $281\ \text{Hz}$ flute it produces $20.0\ \text{Hz}$. What is the frequency of the tuning fork?

Solution

By equation (30.9), the fork's frequency differs from $246\ \text{Hz}$ by $15.0\ \text{Hz}$, so it is either $231\ \text{Hz}$ or $261\ \text{Hz}$. It also differs from $281\ \text{Hz}$ by $20.0\ \text{Hz}$, so it is either $261\ \text{Hz}$ or $301\ \text{Hz}$. The only value consistent with both measurements is

$$f_{\text{fork}} = \boxed{261\ \text{Hz}}. \tag{30.5.1}$$

The two beat measurements each leave a twofold ambiguity, but together they pin the answer down: a small lesson in using two clues to remove an ambiguity that neither could resolve alone.

This page titled 30.5: Beats is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.6: The Doppler Effect

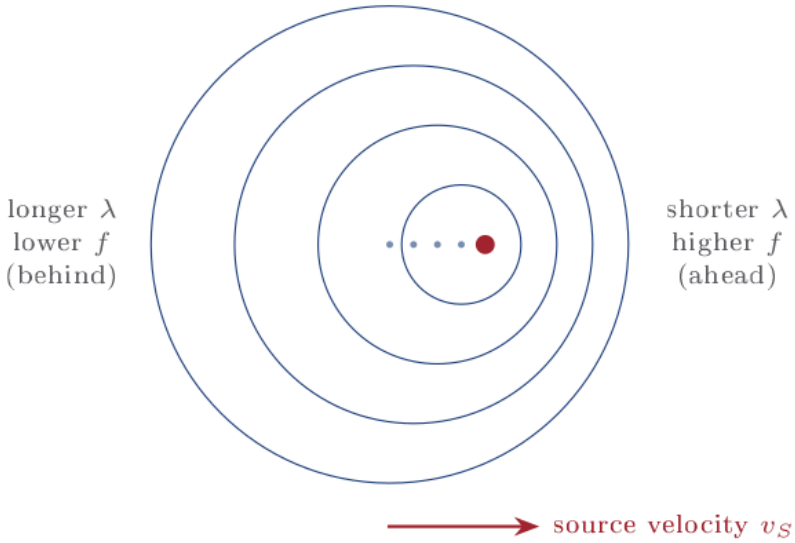


Figure 30.4. The Doppler effect. Each circle is a wavefront, centered on the point (small dot) where the source was when it emitted that front. A source (red) moving to the right crowds its wavefronts together ahead of it (shorter wavelength, higher detected frequency) and stretches them out behind (longer wavelength, lower frequency). An observer ahead hears a higher pitch; one behind, a lower pitch.

•Figure description

A source, drawn as a red dot, moving to the right with speed v sub S, as shown by a red arrow below the drawing. Around it are four circular wavefronts it emitted at equal time intervals. Each circle is centered on where the source was when it emitted that front, so the centers step to the left the further back in time the front was sent out. As a result the circles are crowded close together ahead of the source, on the right, labeled shorter wavelength, higher f, and spread far apart behind it, on the left, labeled longer wavelength, lower f.

We can now explain the ambulance siren. The **Doppler effect** is the change in the frequency an observer detects when there is relative motion between the source of a sound and the observer. When the two approach each other, the wavefronts pile up and the detected frequency rises; when they separate, the wavefronts spread out and the frequency falls (Figure 30.4). The whole effect is captured in one equation:

Key Equation: Doppler effect

$$f' = f\left(\frac{v \pm v_D}{v \pm v_S}\right). \tag{30.10}$$

Here f is the emitted frequency, f' the detected frequency, v the speed of sound in the medium, v_D the speed of the detector, and v_S the speed of the source. The only hard part is the signs, and a single rule fixes them:

> *Toward means up; away means down.* Choose each sign so that motion of the source or detector *toward* the other shifts the frequency *up*, and motion *away* shifts it *down*.

So a detector moving toward the source takes the $+$ sign in the numerator (a larger numerator raises f'), while a source moving toward the detector takes the $-$ sign in the denominator (a smaller denominator raises f'). Reverse each for motion away. Apply the rule to each of v_D and v_S independently, then read off the shift.

Example 30.5: The dropped phone

Your alarm sounds a steady $600\ \text{Hz}$ tone. In frustration you drop the phone, still ringing, out a window $15.0\ \text{m}$ above the ground. Take the speed of sound as $343\ \text{m/s}$.

a. What frequency do you hear from your window just before the phone hits the ground?
b. At what rate is the frequency you hear changing with time at that moment?

Solution

(a) Just before impact the phone has fallen $15.0\ \text{m}$ from rest, so its speed is

$$v_S = \sqrt{2gh} = \sqrt{2(9.8)(15.0)} = 17.1\,\text{m/s}, \tag{30.6.1}$$

directed downward: away from you at the window. The source is moving away, so the frequency shifts *down*; with the detector at rest ($v_D = 0$) and the away sign in the denominator,

$$f' = f\left(\frac{v}{v + v_S}\right) = 600\left(\frac{343}{343 + 17.1}\right) \approx \boxed{571\,\text{Hz}}. \tag{30.6.2}$$

(b) The detected frequency is $f' = fv/(v + v_S)$ with $v_S = gt$ increasing in time. Differentiating,

$$\frac{\mathrm{d}f'}{\mathrm{d}t} = -\frac{f\,v\,g}{(v + v_S)^2} = -\frac{(600)(343)(9.8)}{(343 + 17.1)^2} \approx \boxed{-15.5\ \text{Hz/s}}. \tag{30.6.3}$$

The pitch is sliding downward at about $16\ \text{Hz}$ each second as the phone accelerates away: the falling "wheeee" you have surely heard from a dropped speaker.

Example 30.6: Submarines passing

Submarine A travels at $8.00\ \text{m/s}$ and emits sonar at $1400\ \text{Hz}$; the speed of sound in seawater is $1533\ \text{m/s}$. Submarine B moves directly toward A at $9.00\ \text{m/s}$.

a. What frequency does an observer on B detect as the two approach?
b. After they pass and recede from each other, what frequency does B detect?

Solution

(a) As they approach, both motions are toward the other, so both signs are chosen to shift the frequency up: $+v_D$ in the numerator, $-v_S$ in the denominator.

$$f' = f\left(\frac{v + v_D}{v - v_S}\right) = 1400\left(\frac{1533 + 9.00}{1533 - 8.00}\right) \approx \boxed{1416\,\text{Hz}}. \tag{30.6.4}$$

(b) After passing, both move away from the other, so both signs reverse to shift the frequency down:

$$f' = f\left(\frac{v - v_D}{v + v_S}\right) = 1400\left(\frac{1533 - 9.00}{1533 + 8.00}\right) \approx \boxed{1385\,\text{Hz}}. \tag{30.6.5}$$

The detected pitch jumps from $1416\ \text{Hz}$ to $1385\ \text{Hz}$ as the subs pass: the underwater version of the ambulance siren.

Checkpoint 30.4

An ambulance approaches you at constant speed with its siren on, passes, and drives away. (a) Describe the pitch you hear throughout. In particular, does the pitch keep rising as the ambulance gets closer? (b) What pitch do you hear if you are driving right behind the ambulance at the same speed?

Answer

Answers to Checkpoints

Checkpoint 30.4. (a) While the ambulance approaches at constant speed you hear a steady pitch *higher* than the siren's true

frequency. The pitch does not keep rising as it gets closer; the Doppler shift depends on the source's velocity toward you, not its distance. (It does grow louder.) As it passes, the pitch drops quickly, and as it drives away you hear a steady pitch *lower* than the true one. (b) The true pitch. With no relative motion between source and listener there is no Doppler shift.

This page titled 30.6: The Doppler Effect is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.S: Waves II (Summary)

Summary

- A **sound wave** is a longitudinal wave of alternating compressions and rarefactions. It can be described by displacement $s = A\cos(kx - \omega t)$ or by pressure $\Delta P = \Delta P_{\max}\sin(kx - \omega t)$ with $\Delta P_{\max} = \rho v \omega A$; the two are a quarter-cycle out of phase.
- The **speed of sound** is $v = \sqrt{B/\rho}$ (bulk modulus over density), about $343\ \mathrm{m/s}$ in air: the same "stiffness over inertia" form as the wave speed on a string.
- Two sources in phase interfere according to their **path-length difference**: *constructive* when $\Delta r = m\lambda$, *destructive* when $\Delta r = (m + \frac{1}{2})\lambda$.
- **Intensity** $I = P/A$; for a point source $I = P/4\pi r^2$, an inverse-square law. The **sound level** is $\beta = (10\ \mathrm{dB})\log(I/I_0)$ with $I_0 = 10^{-12}\ \mathrm{W/m^2}$; a factor of 10 in intensity is $+10$ dB, a factor of 2 about $+3$ dB.
- **Beats**: two tones of nearly equal frequency throb at the **beat frequency** $f_{\mathrm{beat}} = |f_2 - f_1|$.
- The **Doppler effect**: $f' = f\dfrac{v \pm v_D}{v \pm v_S}$. Choose signs by the rule *toward* $\rightarrow$ *up, away* $\rightarrow$ *down*, applied to source and detector independently.

Key Equations

Sound wave: displacement and pressure

$$\begin{aligned} s(x,t) &= A\cos(kx - \omega t) \\ \Delta P &= \Delta P_{\max}\sin(kx - \omega t), \qquad \Delta P_{\max} = \rho\, v\, \omega A \end{aligned} \tag{30.S.1}$$

Speed of sound

$$v = \sqrt{\frac{B}{\rho}} \tag{30.S.2}$$

Interference from a path difference

$$\begin{aligned} &\text{constructive:} \quad \Delta r = m\lambda \\ &\text{destructive:} \quad \Delta r = \left(m + \tfrac{1}{2}\right)\lambda \end{aligned} \tag{30.S.3}$$

Intensity

$$I = \frac{P}{A}, \qquad \text{point source: } I = \frac{P}{4\pi r^2} \tag{30.S.4}$$

Comparing two distances

$$\frac{I_2}{I_1} = \left(\frac{r_1}{r_2}\right)^2 \tag{30.S.5}$$

Sound intensity level

$$\beta = (10\ \mathrm{dB})\log\left(\frac{I}{I_0}\right), \qquad I_0 = 10^{-12}\,\mathrm{W/m^2} \tag{30.S.6}$$

Beat frequency

$$f_{\text{beat}} = |f_2 - f_1| \tag{30.S.7}$$

Doppler effect

$$f' = f\left(\frac{v \pm v_D}{v \pm v_S}\right) \quad (\text{toward} \rightarrow \text{up, away} \rightarrow \text{down}) \tag{30.S.8}$$

This page titled 30.S: Waves II (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

30.E: Waves II (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Take the speed of sound in air as $343\ \text{m/s}$ and $I_0 = 10^{-12}\ \text{W/m}^2$ unless told otherwise.

Section 30.1 — Sound Waves

30.1 ★ A sound wave in air has frequency $440\ \text{Hz}$. Find its wavelength.

30.2 ★ *[Conceptual.]* A sound wave is described as either a displacement wave or a pressure wave. Explain why a point of maximum displacement amplitude is a point of zero pressure variation, and vice versa.

Section 30.2 — Speed of Sound

30.3 ★ The bulk modulus of water is about $2.2 \times 10^9\ \text{Pa}$ and its density is $1.0 \times 10^3\ \text{kg/m}^3$. Estimate the speed of sound in water.

30.4 ★★ A sound wave of frequency $250\ \text{Hz}$ travels through a metal rod at $5000\ \text{m/s}$.

a. Find the wavelength in the rod.
b. The same sound passes into air ($v = 343\ \text{m/s}$). Does its frequency change? Find its wavelength in air.

30.5 ★ A sound pulse travels $1.5\ \text{km}$ through air. (a) How long does it take? (b) How long would the same trip take through water, where sound travels at $1480\ \text{m/s}$?

Section 30.3 — Wave Interference

30.6 ★★ Two loudspeakers, driven in phase by the same $686\ \text{Hz}$ source, face a listener. One is $4.00\ \text{m}$ away and the other $4.25\ \text{m}$ away.

a. Find the wavelength of the sound.
b. Find the path-length difference, and state whether the listener hears a maximum or a minimum.

30.7 ★★ Sound from a $500\ \text{Hz}$ speaker splits and recombines through a tunable path, like a trombone slide. Starting from constructive interference, find the minimum slider movement that produces (a) destructive interference and (b) the next constructive interference. (Recall the slider changes the path by twice its displacement.)

30.8 ★★ Two loudspeakers $3.0\ \text{m}$ apart emit identical $686\ \text{Hz}$ tones in phase. A listener stands $4.0\ \text{m}$ directly in front of one of the speakers.

a. Find the difference in path length from the two speakers to the listener.
b. Is the interference there constructive or destructive? Justify your answer.

Section 30.4 — Intensity and Decibels

30.9 ★ A sound has intensity $1.0 \times 10^{-6}\ \text{W/m}^2$. Find its sound level in decibels.

30.10 ★★ A point source radiates $2.5\ \text{W}$ of sound power uniformly in all directions.

a. Find the intensity $5.0\ \text{m}$ from the source.
b. Find the sound level there.
c. At what distance has the level dropped to $40\ \text{dB}$?

30.11 ★★ *[Real-world application: engineering.]* A woman stands a distance d from a loud motor that emits sound uniformly in all directions; the intensity at her position is $3.8 \times 10^{-3}\ \mathrm{W/m^2}$. At a distance 1.8 times as far, find (a) the sound intensity and (b) the sound level relative to the threshold of hearing.

30.12 ★★ A small loudspeaker radiates $2.0\ \mathrm{W}$ of sound power uniformly in all directions. At $5.0\ \mathrm{m}$, find (a) the intensity and (b) the sound level in decibels.

Section 30.5 — Beats

30.13 ★ A piano tuner sounds a $440\ \mathrm{Hz}$ reference fork against a string and hears $3\ \mathrm{Hz}$ beats. What are the two possible frequencies of the string?

30.14 ★★ When a tuning fork is sounded with a $512\ \mathrm{Hz}$ tone, the beat frequency is $4.0\ \mathrm{Hz}$. When the same fork is sounded with a $517\ \mathrm{Hz}$ tone, the beat frequency is $1.0\ \mathrm{Hz}$. Find the frequency of the fork.

30.15 ★ A flute and an oboe play notes of $523\ \mathrm{Hz}$ and $527\ \mathrm{Hz}$ at the same time.

a. What beat frequency do the listeners hear?
b. How much time passes between successive maxima of loudness?
c. The oboist adjusts her reed until the beats slow to one every $2.0\ \mathrm{s}$. What are the two possible new frequencies of the oboe?

Section 30.6 — The Doppler Effect & the Ambulance Siren

30.16 ★ A car horn emits $500\ \mathrm{Hz}$. The car moves toward a stationary listener at $30\ \mathrm{m/s}$. What frequency does the listener hear?

30.17 ★★ *[Real-world application: the ambulance siren.]* An ambulance siren emits $700\ \mathrm{Hz}$. You stand still at the curb.

a. What frequency do you hear as it approaches you at $25\ \mathrm{m/s}$?
b. What frequency do you hear after it passes and recedes at $25\ \mathrm{m/s}$?
c. By how many hertz does the pitch drop as it passes? Explain why a paramedic riding inside hears no change at all.

30.18 ★★ A stock car's engine emits a steady tone. As the car speeds away from you, you detect a frequency 0.847 times the frequency you hear when the car is at rest. Find the speed of the car.

30.19 ★★★ Two trucks travel at the same speed in adjacent lanes, approaching head-on. One driver hears the other's horn at a frequency 1.10 times the frequency heard when both are parked. Take the speed of sound as $344\ \mathrm{m/s}$. Find the common speed of the trucks.

30.20 ★★★ *[Real-world application: biology.]* You fly an ultralight at $40.3\ \mathrm{m/s}$ while an eagle flies directly toward you at $21.4\ \mathrm{m/s}$ (both speeds relative to the ground). The eagle emits a cry at $3830\ \mathrm{Hz}$; the speed of sound is $330\ \mathrm{m/s}$. What frequency do you hear? *(Here both source and detector move, each toward the other.)*

Challenge Problems

30.21 ★★★ *[Real-world application: medicine — Doppler ultrasound.]* A vascular technician aims a $5.0\ \mathrm{MHz}$ ultrasound transducer at an artery, with the beam making an angle $\theta = 60^\circ$ with the direction of blood flow. The speed of sound in soft tissue is $1540\ \mathrm{m/s}$. Only the component of the blood's velocity along the beam, $v\cos\theta$, produces a shift.

a. Red blood cells first *receive* the sound as moving detectors and then *re-emit* it as moving sources. Explain why this makes the observed shift twice what a single application of the Doppler formula would give, so that for $v \ll v_{\text{sound}}$,

$$\Delta f \approx \frac{2 f v \cos\theta}{v_{\text{sound}}}. \tag{30.E.1}$$

b. Blood in the artery flows toward the transducer at $0.30\ \mathrm{m/s}$. Find Δf, and state whether the returned frequency is higher or lower than $5.0\ \mathrm{MHz}$.

c. During systole the instrument reports a shift of $1.30\ \text{kHz}$. Find the peak flow speed.
d. Explain why the technician is trained to avoid angles close to 90°, and what happens to the *fractional* uncertainty in v if θ is misjudged by 5° when the true angle is 80° rather than 60°.

30.22 ★★★ *[Real-world application: engineering — industrial noise control.]* A single machine on a factory floor radiates sound uniformly in all directions and produces a sound level of $85\ \text{dB}$ at a distance of $3.0\ \text{m}$, the limit above which hearing protection is required for an eight-hour shift.

a. Find the sound intensity at $3.0\ \text{m}$, and the acoustic power output of the machine.
b. Eight identical machines are installed close together. Treating their intensities as simply additive, find the sound level at $3.0\ \text{m}$ from the cluster. Note that the level does *not* rise by a factor of eight.
c. How far from the cluster must a worker stand for the level to fall back to $85\ \text{dB}$?
d. Instead of moving the workers, the plant engineer proposes enclosures that reduce each machine's radiated power. By how many decibels must each machine be quieted so that all eight together give $85\ \text{dB}$ at $3.0\ \text{m}$? By what factor must the acoustic power be reduced?

30.23 ★★★ *[Real-world application: music — the concert hall warms up.]* The speed of sound in air depends on absolute temperature as $v = (331\ \text{m/s})\sqrt{T/273\,\text{K}}$. An open–open organ pipe is voiced to sound $A_4 = 440\ \text{Hz}$ as its fundamental when the hall is at 20°C.

a. Find the speed of sound at 20°C from the formula above, confirm it is close to the $343\ \text{m/s}$ used in this chapter, and find the length of the pipe.
b. By the second half of the concert the hall has warmed to 30°C. Find the new speed of sound.
c. The pipe's length is essentially unchanged. Find its new fundamental frequency, and the beat frequency a listener hears when it is sounded together with a piano still tuned to $440\ \text{Hz}$.
d. A piano's steel strings, by contrast, go slightly *flat* as they warm, because the tension drops as the strings expand. Explain why the wind and string sections of an orchestra therefore drift *apart* in pitch as a hall warms, and why players retune at intermission.

This page titled 30.E: Waves II (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31: INTRODUCTION TO OPTICS

Stand a drinking straw in a glass of water and look at it from the side: the straw appears broken at the water's surface, its lower half shifted sideways as though snapped. Put on a pair of polarized sunglasses on a bright day by a lake, and the dazzling glare off the water vanishes while the trees and sky stay clear. Look up from the bottom of a swimming pool and the whole sky seems squeezed into a bright circle directly overhead, surrounded by a silvery mirror. None of these effects needs any new physics beyond what you already know. Each is light, an electromagnetic wave, meeting matter.

Chapter 27 showed that light is an electromagnetic wave, a wave of electric and magnetic fields traveling at c. This chapter begins **optics**, the study of how light travels and how it interacts with mirrors, lenses, and the boundaries between materials. We start by deciding how to picture light: sometimes as a spreading wave, and much more often as a set of straight-line rays. With the ray picture in hand, we work out how light is filtered by polarizers, how it reflects and bends at a surface, how it can be trapped inside a material, and how reflection itself polarizes light. The next chapter uses these rules to form images with mirrors and lenses, and Chapters 33 and 34 return to the wave picture for the effects that rays cannot explain.

Learning Objectives

- Use Huygens' principle to describe how a wave of light propagates, and explain why its wave nature is usually hidden.
- State when the ray approximation is valid, and describe light with rays.
- Apply Malus's law to polarized light and stacks of polarizing filters.
- Apply the laws of reflection and refraction (Snell's law) using the index of refraction.
- Find the critical angle for total internal reflection.
- Find Brewster's angle, at which reflected light is completely polarized.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

31.1: Light as a Wave
31.2: Ray Optics Approximation
31.3: Polarization
31.4: Reflection and Refraction
31.5: Total Internal Reflection
31.6: Polarization by Reflection
31.S: Introduction to Optics (Summary)
31.E: Introduction to Optics (Problems)

This page titled 31: Introduction to Optics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.1: Light as a Wave

Light is an electromagnetic wave. To understand how it travels, bends, and spreads, we need a practical picture of how such a wave moves forward, and the oldest and most useful one is **Huygens' principle**:

> Every point on a wavefront acts as a source of spherical secondary *wavelets*, spreading forward at the wave speed. A short time later, the new wavefront is the surface that just touches (the envelope of) all those wavelets.

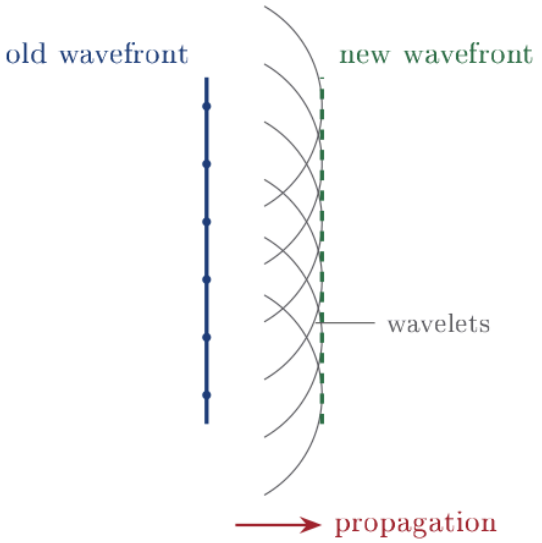


Figure 31.1. Huygens' construction. Each point of the old wavefront emits a spherical wavelet; their common envelope a moment later is the new wavefront. Differential slowing of the wavelets in a medium produces refraction; their sideways spread at an edge produces diffraction.

•Figure description

Huygens' construction. A thick blue vertical line at the left is the old wavefront. Six dots are marked along it, and from each a small arc fans forward to the right -- the secondary wavelets, all of the same radius. A dashed green vertical line to the right just touches the leading edge of every arc: it is the envelope of the wavelets, and it is the new wavefront a moment later. An arrow beneath shows the wave propagating to the right.

A **wavefront** is a surface of constant phase: the crest of a wave, say. For a distant source the wavefronts arriving at a slit are essentially flat, a **plane wave**; near a small source they are spheres. Huygens' construction (Figure 31.1) looks almost too simple to be useful, yet it explains a great deal. It reproduces the law of refraction: when one edge of a wavefront enters glass and slows before the other edge does, the envelope of the lagging wavelets tilts, and the beam bends toward the normal. And it explains **diffraction**: the bending and spreading of waves at apertures and edges. The wavelets at the rim of an opening spread sideways, so a wave does not cast a perfectly sharp shadow; it leaks into the geometric darkness.

This raises a fair question: if waves bend around obstacles, why can you hear someone through an open doorway but not see around it? Both light and sound diffract, but only appreciably when the opening is comparable to the wavelength. Sound wavelengths are a meter or so, about the size of a door, so sound floods through and spreads. Visible light has a wavelength under a thousandth of a millimeter, so a meter-wide door is enormous by comparison and the light marches almost straight through, its diffraction far too slight to notice. To see light behave as a wave, the openings must be shrunk to a hair's width, as Thomas Young did in the experiment of Chapter 33. For almost everything else, light's wave nature stays hidden, and the next section takes advantage of that.

Checkpoint 31.1

A radio station broadcasts at $1.0\ \text{MHz}$ (wavelength about $300\ \text{m}$); a tomato-red laser emits at $650\ \text{nm}$. Which wave diffracts more strongly around a building, and why does this help explain why radio reaches you in a tunnel but a flashlight beam does not bend around the corner?

Answer

Answers to Checkpoints

Checkpoint 31.1. The radio wave. Diffraction spreads a wave around an obstacle through an angle of about λ/a, and it is strong only when the wavelength is comparable to or larger than the obstacle. A $300\ \mathrm{m}$ radio wave is as large as a building and bends readily around it and into a tunnel's mouth. Light's wavelength, $650\ \mathrm{nm}$, is about a ten-millionth the size of a tunnel entrance, so it hardly diffracts at all: it travels in straight lines and leaves a sharp shadow.

This page titled 31.1: Light as a Wave is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.2: Ray Optics Approximation

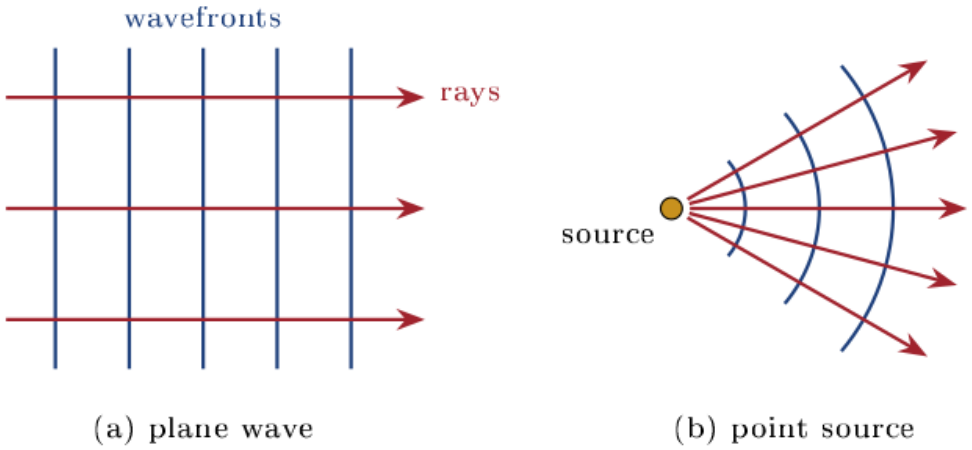


Figure 31.2. Wavefronts (blue) and rays (red). Rays are drawn perpendicular to the wavefronts, in the direction the wave travels. (a) A plane wave has parallel rays. (b) Spherical waves from a small source have rays that radiate outward from it.

•Figure description

Two panels showing wavefronts and rays. In panel a, a plane wave is drawn as a set of parallel, evenly spaced vertical lines, the wavefronts, with horizontal arrows, the rays, crossing them at right angles and pointing to the right. In panel b, a small point source at the left sends out circular arcs, the spherical wavefronts, with straight arrows radiating outward from the source, the rays, crossing every arc at right angles.

A wavefront is a surface of constant phase, and the direction in which the wave travels is perpendicular to it everywhere. A line drawn in that direction, always perpendicular to the wavefronts, is called a **ray**. For a plane wave the rays are parallel straight lines. For the spherical wavefronts spreading from a small source they are straight lines radiating outward from the source (Figure 31.2). A ray is not a physical object. It is a way of summarizing where the wave's energy goes.

The **ray approximation** is the assumption that light travels along straight rays in any uniform medium, changing direction only where it meets a surface, where it reflects or refracts. The study of light in this approximation is called **ray optics**, or geometric optics, because it reduces optics to geometry: drawing straight lines and measuring angles.

When is the approximation valid? The previous section gave the answer. A wave passing through an opening or past an edge spreads sideways, diffracting into the geometric shadow, but only by an angle of roughly

$$\theta \sim \frac{\lambda}{a}, \tag{31.1}$$

where a is the width of the opening. Chapter 34 makes this precise. When the openings and objects the light meets are much larger than its wavelength, $a \gg \lambda$, the spreading is negligible, the light travels in straight lines, and objects cast sharp shadows. *Ray optics is the limit of wave optics when the wavelength is small compared with everything the light encounters.* Visible light has wavelengths below a micrometer, so for mirrors, lenses, windows, and eyes, all millimeters to meters in size, the ray approximation is excellent, and it is the model used in the rest of this chapter and in Chapter 32. When light passes through openings comparable to its wavelength, such as narrow slits, fine gratings, and tiny apertures, rays fail and the wave nature takes over. Those effects are the subject of Chapters 33 and 34.

Example 31.1: How well does a laser beam stay straight?

A red laser beam of wavelength $633\ \text{nm}$ leaves the laser through a circular opening $2.0\ \text{mm}$ across. Estimate the angle through which the beam spreads by diffraction, and how much wider it is after traveling $100\ \text{m}$.

Solution

From equation (31.1),

$$\theta \sim \frac{\lambda}{a} = \frac{633 \times 10^{-9}\,\mathrm{m}}{2.0 \times 10^{-3}\,\mathrm{m}} \approx \boxed{3 \times 10^{-4}\,\mathrm{rad}}, \tag{31.2.1}$$

about 0.02°. Over $100\ \mathrm{m}$ the beam's edges spread by roughly $(100\ \mathrm{m})(3 \times 10^{-4}) = 3\ \mathrm{cm}$ on each side, so the spot is several centimeters across. That is small compared with the distance traveled, which is why a laser beam looks like a straight line, the ideal of a ray. But it is not zero, and no beam of finite width can be perfectly parallel. The wave nature of light is always there. Ray optics simply ignores it when it is small.

Checkpoint 31.2

A pinhole camera is a light-tight box with a tiny hole in one side and film (or a sensor) on the opposite side. (a) Using rays, explain why a *smaller* pinhole gives a sharper image. (b) Using equation (31.1), explain why making the pinhole smaller and smaller eventually makes the image *blurrier* again. (c) Where does the ray approximation break down in this camera?

Answer

Answers to Checkpoints

Checkpoint 31.2. (a) Each point of the object sends light through the pinhole to a small patch on the film, the size of the hole. A smaller hole shrinks every patch, so neighboring points overlap less and the image is sharper. (b) The light spreads by diffraction through an angle $\theta \sim \lambda/a$, which *grows* as the hole width a shrinks. Below some size the diffraction spread on the film outweighs the geometric patch, and the image blurs again. (c) The ray approximation fails when the hole is no longer much larger than the scale set by diffraction, when the spread $L\lambda/a$ across the box of depth L becomes comparable to the hole size a itself. The sharpest image comes from a pinhole near that crossover, $a \sim \sqrt{L\lambda}$, a fraction of a millimeter for a typical box.

This page titled 31.2: Ray Optics Approximation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.3: Polarization

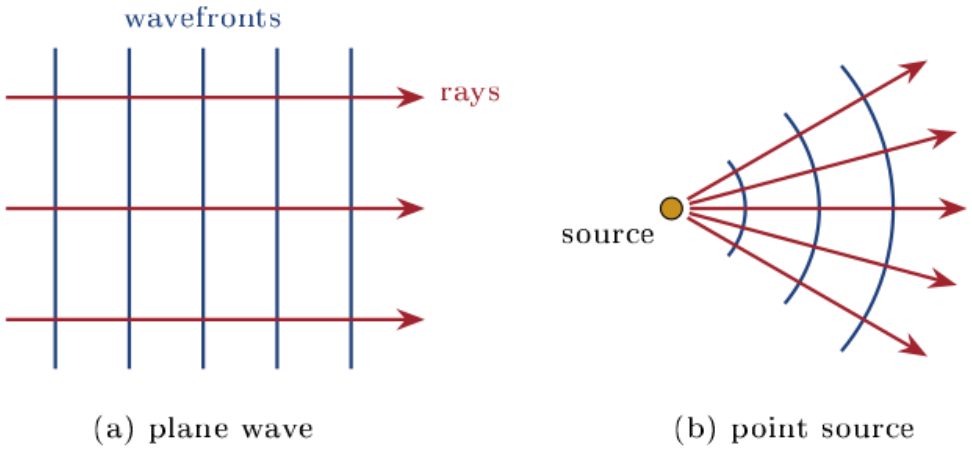


Figure 31.2. Wavefronts (blue) and rays (red). Rays are drawn perpendicular to the wavefronts, in the direction the wave travels. (a) A plane wave has parallel rays. (b) Spherical waves from a small source have rays that radiate outward from it.

•Figure description

Two panels showing wavefronts and rays. In panel a, a plane wave is drawn as a set of parallel, evenly spaced vertical lines, the wavefronts, with horizontal arrows, the rays, crossing them at right angles and pointing to the right. In panel b, a small point source at the left sends out circular arcs, the spherical wavefronts, with straight arrows radiating outward from the source, the rays, crossing every arc at right angles.

A wavefront is a surface of constant phase, and the direction in which the wave travels is perpendicular to it everywhere. A line drawn in that direction, always perpendicular to the wavefronts, is called a **ray**. For a plane wave the rays are parallel straight lines. For the spherical wavefronts spreading from a small source they are straight lines radiating outward from the source (Figure 31.2). A ray is not a physical object. It is a way of summarizing where the wave's energy goes.

The **ray approximation** is the assumption that light travels along straight rays in any uniform medium, changing direction only where it meets a surface, where it reflects or refracts. The study of light in this approximation is called **ray optics**, or geometric optics, because it reduces optics to geometry: drawing straight lines and measuring angles.

When is the approximation valid? The previous section gave the answer. A wave passing through an opening or past an edge spreads sideways, diffracting into the geometric shadow, but only by an angle of roughly

$$\theta \sim \frac{\lambda}{a}, \tag{31.1}$$

where a is the width of the opening. Chapter 34 makes this precise. When the openings and objects the light meets are much larger than its wavelength, $a \gg \lambda$, the spreading is negligible, the light travels in straight lines, and objects cast sharp shadows. *Ray optics is the limit of wave optics when the wavelength is small compared with everything the light encounters.* Visible light has wavelengths below a micrometer, so for mirrors, lenses, windows, and eyes, all millimeters to meters in size, the ray approximation is excellent, and it is the model used in the rest of this chapter and in Chapter 32. When light passes through openings comparable to its wavelength, such as narrow slits, fine gratings, and tiny apertures, rays fail and the wave nature takes over. Those effects are the subject of Chapters 33 and 34.

Example 31.1: How well does a laser beam stay straight?

A red laser beam of wavelength $633\ \text{nm}$ leaves the laser through a circular opening $2.0\ \text{mm}$ across. Estimate the angle through which the beam spreads by diffraction, and how much wider it is after traveling $100\ \text{m}$.

Solution

From equation (31.1),

$$\theta \sim \frac{\lambda}{a} = \frac{633 \times 10^{-9}\,\mathrm{m}}{2.0 \times 10^{-3}\,\mathrm{m}} \approx \boxed{3 \times 10^{-4}\,\mathrm{rad}}, \tag{31.3.1}$$

about 0.02°. Over $100\ \mathrm{m}$ the beam's edges spread by roughly $(100\ \mathrm{m})(3 \times 10^{-4}) = 3\ \mathrm{cm}$ on each side, so the spot is several centimeters across. That is small compared with the distance traveled, which is why a laser beam looks like a straight line, the ideal of a ray. But it is not zero, and no beam of finite width can be perfectly parallel. The wave nature of light is always there. Ray optics simply ignores it when it is small.

Checkpoint 31.2

A pinhole camera is a light-tight box with a tiny hole in one side and film (or a sensor) on the opposite side. (a) Using rays, explain why a *smaller* pinhole gives a sharper image. (b) Using equation (31.1), explain why making the pinhole smaller and smaller eventually makes the image *blurrier* again. (c) Where does the ray approximation break down in this camera?

Answer

Answers to Checkpoints

Checkpoint 31.2. (a) Each point of the object sends light through the pinhole to a small patch on the film, the size of the hole. A smaller hole shrinks every patch, so neighboring points overlap less and the image is sharper. (b) The light spreads by diffraction through an angle $\theta \sim \lambda/a$, which *grows* as the hole width a shrinks. Below some size the diffraction spread on the film outweighs the geometric patch, and the image blurs again. (c) The ray approximation fails when the hole is no longer much larger than the scale set by diffraction, when the spread $L\lambda/a$ across the box of depth L becomes comparable to the hole size a itself. The sharpest image comes from a pinhole near that crossover, $a \sim \sqrt{L\lambda}$, a fraction of a millimeter for a typical box.

This page titled 31.3: Polarization is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.4: Reflection and Refraction

When light meets a boundary between two transparent media, part of it bounces back (reflection) and part crosses into the new medium, generally bending (refraction). We describe both with rays, in the ray approximation of Section 31.2.

Measure all angles from the **normal**, the line perpendicular to the surface. The **law of reflection** is then beautifully simple: the angle of reflection equals the angle of incidence,

Key Equation: Law of reflection

$$\theta_i = \theta_r. \tag{31.4}$$

This single rule explains both a mirror and a sheet of paper. A mirror is smooth on the scale of a wavelength, so parallel incoming rays stay parallel after reflecting (*specular* reflection) and an image forms. Paper is rough on that scale, so its tiny facets scatter parallel rays in all directions (*diffuse* reflection); each individual ray still obeys $\theta_i = \theta_r$, but the surface as a whole throws the light everywhere and you see no image.

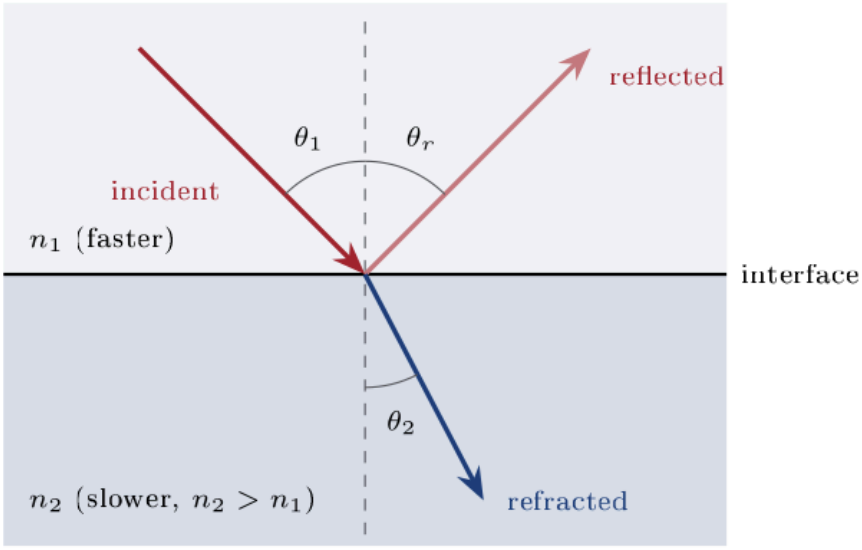


Figure 31.4. Reflection and refraction at a boundary. The reflected ray obeys $\theta_r = \theta_1$. The refracted ray bends toward the normal on entering the slower medium ($n_2 > n_1$), as Snell's law $n_1 \sin\theta_1 = n_2 \sin\theta_2$ requires.

•Figure description

A ray diagram at a horizontal interface between two media, the lighter upper medium n sub 1 being faster and the darker lower medium n sub 2 slower. A dashed vertical normal line crosses the interface at the point where a red incident ray arrives from the upper left at angle theta sub 1 to the normal. A reflected ray leaves back into the upper medium at an equal angle theta sub r on the other side of the normal. A blue refracted ray continues into the lower medium at a smaller angle theta sub 2, bent toward the normal because the lower medium is slower. All three angles are marked with arcs.

Refraction requires one new idea. Light travels at c only in vacuum; in a material medium with permittivity ϵ and permeability μ it slows to $v = 1/\sqrt{\mu\epsilon} < c$. We capture this slowdown in a single number, the **index of refraction**:

$$n = \frac{c}{v} \quad (\geq 1). \tag{31.5}$$

Vacuum has $n = 1$; air is very nearly 1; water is 1.33; typical glass about 1.52; diamond 2.42. The greater n is, the slower light moves and the more sharply a ray bends. The bending is governed by **Snell's law**, which relates the angles on the two sides of the boundary to the two indices (Figure 31.4):

Key Equation: Snell's law of refraction

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{31.6}$$

A ray entering a *slower* (higher-n) medium bends *toward* the normal; entering a faster (lower-n) medium it bends away. (The frequency of the light is unchanged on crossing; since $v = f\lambda$ changes, the wavelength in the medium shrinks to $\lambda_n = \lambda/n$.)

Example 31.3: Into a slab of crown glass

A ray of wavelength $589\ \text{nm}$ traveling through air ($n_1 = 1.00$) strikes a flat slab of crown glass ($n_2 = 1.52$) at 30.0° to the normal.

a. Find the angle of refraction.
b. Find the speed of the light once it enters the glass.

Solution

(a) Apply Snell's law, equation (31.6), and solve for θ_2:

$$\sin\theta_2 = \frac{n_1 \sin\theta_1}{n_2} = \frac{(1.00)\sin 30.0^\circ}{1.52} = \frac{0.500}{1.52} = 0.329, \tag{31.4.1}$$

$$\theta_2 = \arcsin(0.329) = \boxed{19.2^\circ}. \tag{31.4.2}$$

The ray bends toward the normal, as expected on entering a slower medium.

(b) The speed in the glass follows from the definition of the index, equation (31.5):

$$v = \frac{c}{n_2} = \frac{3.00 \times 10^8}{1.52} = \boxed{1.97 \times 10^8\ \text{m/s}}. \tag{31.4.3}$$

Light loses about a third of its vacuum speed inside the glass, and it is precisely this slowing that bends the ray.

Example 31.4: A ray through a parallel-sided slab

A ray passes from medium 1 into a thick slab of index n_2 with flat, parallel faces, and exits back into medium 1 on the far side. Show that the emerging ray is parallel to the original incident ray.

Solution

At the first face Snell's law gives

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{31.4.4}$$

Because the two faces are parallel, the normal at the second face is parallel to the normal at the first, and the ray inside the slab strikes the second face at the same angle θ_2 it made leaving the first. Applying Snell's law at the second face, with the ray going from n_2 back into n_1 and emerging at angle θ_3,

$$n_2 \sin\theta_2 = n_1 \sin\theta_3. \tag{31.4.5}$$

The right-hand side of the first equation equals the left-hand side of the second, so $n_1 \sin\theta_1 = n_1 \sin\theta_3$, giving

$$\boxed{\theta_3 = \theta_1}. \tag{31.4.6}$$

The exit ray makes the same angle with the normal as the entry ray, so the two are parallel. The slab does not change the ray's *direction*; it merely shifts it sideways by a small parallel displacement. (This is why looking through a thick window does not visibly distort the scene: it only nudges it.)

Checkpoint 31.4

You see your face in a bathroom mirror but not in a sheet of white printer paper, even though both are reflecting nearly all the light that hits them. Using the law of reflection, explain the difference between the two surfaces.

Answer

Answers to Checkpoints

Checkpoint 31.4. Both surfaces obey the law of reflection at every point; the difference is their smoothness. A mirror is flat on the scale of the light's wavelength, so parallel rays arriving from your face leave parallel and preserve the image: specular reflection. Paper is a jumble of fibers tilted every which way, so each ray reflects off a surface element with its own orientation, and the rays scatter in all directions: diffuse reflection. The paper returns the light, but not the image.

This page titled 31.4: Reflection and Refraction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.5: Total Internal Reflection

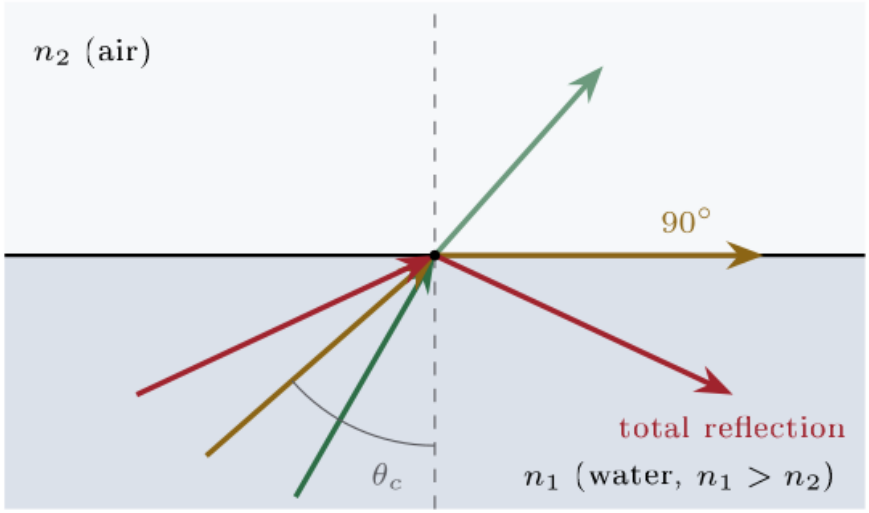


Figure 31.5. Total internal reflection. Below the critical angle (green) light refracts out into the faster medium. At θ_c (gold) the refracted ray grazes the surface at 90°. Beyond θ_c (red) no light escapes: it is entirely reflected back.

•Figure description

A ray diagram at a horizontal water-to-air interface, with the denser water n sub 1 below and air n sub 2 above, and a dashed vertical normal. Three rays travel up through the water to the same point on the surface at increasing angles from the normal. The green ray, at the smallest angle, refracts out into the air. The gold ray arrives at the critical angle theta sub c, marked with an arc, and its refracted ray grazes along the surface at 90 degrees. The red ray, at a larger angle still, has no refracted ray at all -- it reflects entirely back down into the water, labeled total reflection.

Something dramatic happens when light tries to leave a slower medium for a faster one: say, water for air, or glass for air. Going from high n_1 to low n_2, Snell's law bends the refracted ray *away* from the normal, so $\theta_2 > \theta_1$. As you increase the angle of incidence, the refracted ray swings closer and closer to the surface until, at one special incidence angle, it grazes along the boundary at exactly $\theta_2 = 90^\circ$. Beyond that angle there is no refracted ray at all: the light is completely reflected back into the original medium. This is **total internal reflection**, and the threshold angle is the **critical angle** θ_c (Figure 31.5).

Set $\theta_2 = 90^\circ$ in Snell's law: $n_1 \sin\theta_c = n_2 \sin 90^\circ = n_2$, so

Key Equation: Critical angle

$$\theta_c = \arcsin\frac{n_2}{n_1}, \qquad (n_1 > n_2). \tag{31.7}$$

Total internal reflection only occurs when light is in the *denser* medium ($n_1 > n_2$); otherwise the ratio n_2/n_1 exceeds 1 and the arcsine has no solution. Far from a curiosity, this is the principle that traps a light pulse inside an optical fiber, bouncing it down hundreds of kilometers of glass to carry the internet, and it is what makes a cut diamond flash: light entering the gem is repeatedly trapped and reflected by its facets before escaping toward your eye.

Example 31.5: Critical angle for water

Find the critical angle for an air–water boundary. (Take the index of refraction of water to be 1.33.)

Solution

The light is inside the water (the denser medium), so $n_1 = 1.33$ and the second medium is air, $n_2 = 1.00$. Equation (31.7) gives

$$\theta_c = \arcsin\frac{n_2}{n_1} = \arcsin\frac{1.00}{1.33} = \arcsin(0.752) = \boxed{48.8^\circ}. \tag{31.5.1}$$

Any ray striking the underside of the water's surface at more than 48.8° from the normal cannot get out: it is reflected back down. This is why, looking up from underwater, the bright world above is squeezed into a circular "window" directly overhead, ringed by a mirror-like surface reflecting the bottom.

Checkpoint 31.5

Diamond ($n = 2.42$) has a much smaller critical angle than glass ($n = 1.52$). Without computing, predict which gemstone traps light through more internal bounces before it escapes, and explain why this contributes to a diamond's brilliance.

Answer

Answers to Checkpoints

Checkpoint 31.5. Diamond. A smaller critical angle (24° for diamond, about 41° for glass) means a larger range of angles at which light striking the inside of a facet is totally reflected. Light entering a well-cut diamond bounces around inside several times before it meets a facet nearly head-on enough to escape, and it tends to leave through the top toward the viewer. This makes the stone sparkle, and the long path also spreads the colors apart, adding the "fire."

This page titled 31.5: Total Internal Reflection is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.6: Polarization by Reflection

Light reflected from a nonmetallic surface, such as water, glass, or a road, is generally partly polarized, and at one particular angle of incidence it is *completely* polarized. That is why polarized sunglasses can cut glare. To see where the polarization comes from, think about what produces the reflected light. The incoming wave sets the electrons in the surface of the material oscillating along the direction of its electric field, and those oscillating charges radiate the reflected and refracted waves. An oscillating charge radiates most strongly perpendicular to its motion and not at all along the line of its motion.

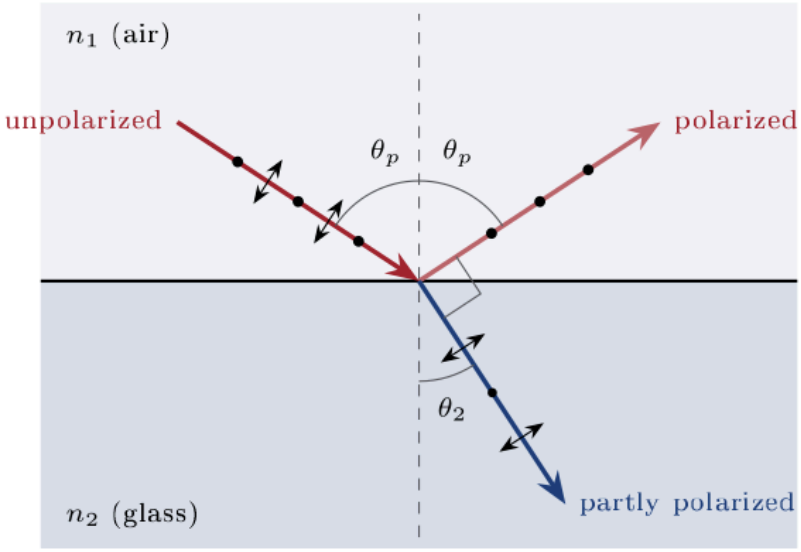


Figure 31.6. Polarization by reflection at Brewster's angle. The dots represent the electric-field component perpendicular to the page and the double arrows the component in the plane of the page. When the reflected and refracted rays are perpendicular, the in-plane component cannot be reflected, and the reflected light is completely polarized parallel to the surface.

•Figure description

A ray diagram at a horizontal interface, with air, index n sub 1, above and water or glass, index n sub 2, below, and a dashed vertical normal. Unpolarized light arrives from the upper left at Brewster's angle theta sub p, drawn with both dots and short double-headed arrows along it to show both polarization components. The reflected ray leaves to the upper right at the same angle and carries only dots, meaning it is polarized perpendicular to the page, parallel to the surface. The refracted ray continues into the lower medium at angle theta sub 2 and carries mostly arrows with a few dots. A small square marks the ninety-degree angle between the reflected and refracted rays.

Split the incident light's electric field into two components (Figure 31.6): one perpendicular to the *plane of incidence* (the plane containing the incident ray and the normal), drawn as dots, and one lying in that plane, drawn as short double-headed arrows. Electrons driven by the perpendicular component oscillate perpendicular to the plane of incidence and can radiate in any direction within it, so that component always reflects. Electrons driven by the in-plane component oscillate perpendicular to the *refracted* ray. If the reflected ray happens to point along that same direction of oscillation, those electrons cannot radiate into it, and the in-plane component is missing from the reflected light entirely. The reflected light is then completely polarized perpendicular to the plane of incidence.

This happens when the reflected and refracted rays are perpendicular to each other. The angle of incidence at which it occurs is called the **polarizing angle** or **Brewster's angle** θ_p. With the reflected ray at θ_p from the normal and the refracted ray at θ_2, the condition that the two rays are perpendicular is $\theta_p + 90^\circ + \theta_2 = 180^\circ$, or $\theta_2 = 90^\circ - \theta_p$. Snell's law then gives

$$n_1 \sin\theta_p = n_2 \sin(90^\circ - \theta_p) = n_2 \cos\theta_p, \tag{31.6.1}$$

and so

Key Equation: Brewster's angle

$$\tan\theta_p = \frac{n_2}{n_1}. \tag{31.8}$$

At Brewster's angle the reflected light is completely polarized, parallel to the surface. The refracted light carries all of the in-plane component and part of the perpendicular one, so it is only partly polarized. At other angles the reflected light is partly polarized, still mostly parallel to the surface.

Example 31.6: Brewster's angle for water and glass

Find Brewster's angle for light in air reflecting from (a) water ($n = 1.33$) and (b) crown glass ($n = 1.52$). (c) For water, find the angle of refraction and check that the reflected and refracted rays are perpendicular.

Solution

(a) From equation (31.8) with $n_1 = 1.00$,

$$\theta_p = \arctan\frac{1.33}{1.00} = \boxed{53.1^\circ}. \tag{31.6.2}$$

(b) For glass, $\theta_p = \arctan 1.52 = \boxed{56.7^\circ}$.

(c) Snell's law gives $\sin\theta_2 = \sin 53.1^\circ/1.33 = 0.600$, so $\theta_2 = \boxed{36.9^\circ}$. The reflected ray is at 53.1° from the normal on one side and the refracted ray at 36.9° from it on the other, so the angle between them is $180^\circ - 53.1^\circ - 36.9^\circ = 90.0^\circ$, as it must be.

Note: Polarized sunglasses

Sunlight reflecting off a lake, a wet road, or a car's hood strikes these horizontal surfaces at fairly large angles of incidence, often near Brewster's angle, so the glare is strongly polarized horizontally. Polarized sunglasses are simply polarizing filters with their transmission axes vertical. By Malus's law they block most of the horizontally polarized glare while still passing about half of the unpolarized light from everything else. Photographers use rotatable polarizing filters on their lenses for the same reason, to cut reflections from water and glass or to darken a blue sky, whose scattered light is also partly polarized. You can check whether a pair of sunglasses is polarized by tilting your head while looking at glare: if the glare brightens as you tilt, it is.

Checkpoint 31.6

(a) You look at light reflected from a pond through a polarizing filter, and rotate the filter. How can you tell whether the light is reaching you at Brewster's angle? (b) Should the transmission axis of polarized sunglasses be vertical or horizontal? What would happen to the glare if you wore them sideways?

Answer

Answers to Checkpoints

Checkpoint 31.6. (a) At Brewster's angle the reflected light is completely polarized, parallel to the water's surface. Rotating the filter, you would find one orientation at which the reflection vanishes entirely. At other angles of reflection the light is only partly polarized, and the reflection merely dims to a minimum. (b) Vertical. Glare from horizontal surfaces such as water and roads is mostly polarized horizontally, so a vertical transmission axis blocks it. Worn sideways, the glasses' axis would be horizontal: they would pass the glare while blocking vertically polarized light, making the glare relatively *worse*.

This page titled 31.6: Polarization by Reflection is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.S: Introduction to Optics (Summary)

Summary

- **Huygens' principle**: every point on a wavefront emits spherical wavelets whose envelope is the next wavefront. It explains refraction and **diffraction**. Waves spread noticeably only through openings comparable to their wavelength.
- A **ray** is drawn perpendicular to the wavefronts, in the direction the wave travels. The **ray approximation** treats light as traveling in straight lines, changing direction only at surfaces. It is valid when every opening and object is much larger than the wavelength, since diffraction spreads light only by an angle of order λ/a.
- **Polarization**: unpolarized light loses half its intensity through one polarizer ($I = \frac{1}{2}I_0$); already-polarized light obeys **Malus's law** $I = I_0 \cos^2\theta$.
- **Reflection**: $\theta_i = \theta_r$. **Refraction**: $n = c/v$ and Snell's law $n_1 \sin\theta_1 = n_2 \sin\theta_2$; rays bend toward the normal entering a slower (higher-n) medium.
- **Total internal reflection** occurs for light leaving a denser medium ($n_1 > n_2$) above the critical angle $\theta_c = \arcsin(n_2/n_1)$. It is the principle behind optical fibers and a diamond's sparkle.
- **Polarization by reflection**: at Brewster's angle, $\tan\theta_p = n_2/n_1$, the reflected and refracted rays are perpendicular and the reflected light is completely polarized parallel to the surface, which is why polarized sunglasses cut glare.

Key Equations

Spreading of a beam by diffraction

$$\theta \sim \frac{\lambda}{a} \tag{31.S.1}$$

Unpolarized light through one polarizer

$$I = \tfrac{1}{2}I_0 \tag{31.S.2}$$

Malus's law

$$I = I_0 \cos^2\theta \tag{31.S.3}$$

Law of reflection

$$\theta_i = \theta_r \tag{31.S.4}$$

Index of refraction

$$n = \frac{c}{v} \quad (\geq 1), \qquad \lambda_n = \frac{\lambda}{n} \tag{31.S.5}$$

Snell's law of refraction

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{31.S.6}$$

Critical angle

$$\theta_c = \arcsin\frac{n_2}{n_1}, \qquad (n_1 > n_2) \tag{31.S.7}$$

Brewster's angle

$$\tan\theta_p = \frac{n_2}{n_1} \tag{31.S.8}$$

This page titled 31.S: Introduction to Optics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

31.E: Introduction to Optics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $c = 3.00 \times 10^8\ \text{m/s}$, $n_{\text{air}} = 1.00$, $n_{\text{water}} = 1.33$, and $n_{\text{glass}} = 1.52$.

Section 31.1 — Light as a Wave

31.1 ★ *[Conceptual.]* State Huygens' principle in your own words, and use it to explain why a wave spreads out after passing through a narrow opening.

31.2 ★ A beam of $600\ \text{nm}$ light passes from air into water ($n = 1.33$). In the water, find its (a) speed, (b) wavelength, and (c) frequency.

Section 31.2 — Ray Optics Approximation

31.3 ★ *[Conceptual.]* A student says: "Light travels in straight lines, so light is made of rays." Explain what is right and what is wrong with this statement.

31.4 ★★ *[Real-world application: astronomy.]* A laser beam of wavelength $532\ \text{nm}$ is aimed at the Moon, $3.84 \times 10^8\ \text{m}$ away, through a telescope whose opening is $1.0\ \text{m}$ across. Estimate the diameter of the spot the beam makes on the Moon. Is the ray approximation adequate for this beam?

Section 31.3 — Polarization

31.5 ★ Polarized light of intensity I_0 strikes a polarizer whose axis is 25° from the light's polarization. What fraction is transmitted?

31.6 ★★ Unpolarized light of intensity I_0 passes through two ideal polarizers whose axes are 40° apart. What fraction of the original light emerges?

31.7 ★★★ Unpolarized light passes through three ideal polarizers. The first is vertical, the second at 30° to vertical, the third at 60° to vertical. What fraction of the incident light emerges from the stack?

Sections 31.4–31.5 — Reflection, Refraction, and Total Internal Reflection

31.8 ★ A light ray strikes a flat mirror at 35° from the normal. What is the angle of reflection?

31.9 ★★ A ray of light in air enters water ($n = 1.33$) at an angle of incidence of 45°.

a. Find the angle of refraction.
b. Find the speed of the light in the water.

31.10 ★★ Find the critical angle for light passing from crown glass ($n = 1.52$) into air.

31.11 ★★★ *[Real-world application: optics.]* A ray inside a glass block ($n = 1.52$) strikes the glass–air boundary.

a. Find the critical angle.
b. A second block is placed against the first with a thin film of water ($n = 1.33$) between them. Find the new critical angle for the glass–water boundary, and explain why a fiber-optic cable's outer cladding is chosen to have an index only slightly below the core's.

Section 31.6 — Polarization by Reflection

31.12 ★ Find Brewster's angle for light in air reflecting from diamond ($n = 2.42$).

31.13 ★★ Light traveling in water ($n = 1.33$) reflects from the surface of a submerged glass plate ($n = 1.52$). Find Brewster's angle, and the angle of refraction in the glass at that angle of incidence.

31.14 ★★ Light in air strikes a transparent material at Brewster's angle and is refracted at 35.0° from the normal. Find the index of refraction of the material.

31.15 ★★ *[Conceptual.]* Unpolarized light is reflected from a glass surface at Brewster's angle and then passes through a polarizing filter. Describe how the transmitted intensity changes as the filter is rotated through a full turn.

Challenge Problems

31.16 ★★★ *[Real-world application: telecommunications — optical fiber.]* A step-index optical fiber has a core of index $n_1 = 1.482$ surrounded by cladding of index $n_2 = 1.465$. Light is guided by total internal reflection at the core–cladding boundary.

a. Find the critical angle at the core–cladding boundary.
b. Light enters the flat end of the fiber from air. Show that a ray is guided only if its angle of incidence at the entrance face satisfies $\sin\theta_a \le \sqrt{n_1^2 - n_2^2}$ (this quantity is the fiber's *numerical aperture*), and evaluate θ_a.
c. A ray traveling straight down the axis covers a length L of fiber in a time $n_1 L/c$. A ray bouncing at the critical angle travels a longer path. Show that the spread in arrival times over a length L is

$$\Delta t = \frac{n_1 L}{c}\left(\frac{1}{\sin\theta_c} - 1\right), \tag{31.E.1}$$

and evaluate it for $L = 1.0\ \text{km}$.

d. Two pulses can be told apart only if they are separated by more than about $2\ \Delta t$. Estimate the maximum bit rate this fiber can carry over $1.0\ \text{km}$, and explain why long-haul links use *single-mode* fiber, whose core is only a few micrometers across, instead.

31.17 ★★★ *[Real-world application: optics — rotating a beam's polarization.]* Polarized light of intensity I_0 must be rotated through 90° without much loss. A single polarizer set at 90° to the incoming polarization transmits nothing. Instead, a stack of N ideal polarizers is used, each rotated by $90^\circ/N$ from the one before it.

a. Show that the fraction transmitted through the whole stack is

$$\frac{I}{I_0} = \cos^{2N}\!\left(\frac{90^\circ}{N}\right). \tag{31.E.2}$$

b. Evaluate this for $N = 1$, $N = 2$, $N = 3$, and $N = 10$.
c. For large N, use $\cos x \approx 1 - x^2/2$ (with $x = \pi/2N$ in radians) and $(1-u)^{2N} \approx e^{-2Nu}$ to show that $I/I_0 \approx e^{-\pi^2/4N}$. Check this against your exact $N = 10$ result, and state the limit as $N \to \infty$.
d. Comment on the trade being made: nothing here violates the fact that a single 90° polarizer blocks everything. Explain in one or two sentences what each small rotation does to the light, and why real devices that rotate polarization (such as the twisted liquid-crystal cell in an LCD) are built as a smooth gradual twist rather than an abrupt one.

This page titled 31.E: Introduction to Optics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32: IMAGES

Stand in front of a bathroom mirror and raise your right hand. The figure facing you raises its left. Read the word ambulance stenciled in reverse across the front of an emergency van and your mirror untangles it instantly. Yet here is a puzzle that has bothered thoughtful people since antiquity: a mirror flips left and right, but it does *not* flip up and down. Your reflection's head stays up and its feet stay down. Why should a flat sheet of glass care about the difference between horizontal and vertical? The mirror, of course, does nothing of the kind, and by the end of this chapter you will be able to say exactly what it does instead.

Answering that question is really the business of this whole chapter: understanding *images*, where light decides to reassemble a picture of an object after bouncing off a mirror or passing through a lens. The same handful of ideas that explain your reflection also explain how a telescope gathers a galaxy, how a camera freezes a face, how your own eye throws the world upside down onto your retina, and how a magnifying glass enlarges the fine print. We will build the toolkit one optic at a time (flat mirrors, then curved ones, then lenses, then whole systems of them) and we will return at the end to the mirror that started it all.

Learning Objectives

- Distinguish real from virtual images and explain how each is formed.
- Locate and describe the image in a plane mirror, and explain the left–right "reversal."
- Apply the mirror equation and magnification to concave and convex spherical mirrors, using a consistent sign convention.
- Trace principal rays to locate images for mirrors and lenses.
- Apply the thin-lens equation to converging and diverging lenses, and use the lensmaker's equation to relate focal length to a lens's shape.
- Analyze multi-element optical systems by treating each image as the object for the next optic.
- Explain how the eye focuses, correct near- and farsightedness with lenses of the right power, and find the magnification of a magnifier, a microscope, and a telescope.

We will return to the left–right puzzle of the plane mirror in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

32.1: Images and Plane Mirrors
32.2: Spherical Mirrors
32.3: Thin Lenses
32.4: Optical Systems
32.5: Optical Instruments
32.S: Images (Summary)
32.E: Images (Problems)

This page titled 32: Images is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.1: Images and Plane Mirrors

An **image** is a reproduction of an object formed by light. It appears wherever the rays of light leaving a point on the object are gathered back to a single point, either because they truly meet there, or because they only appear to. That distinction sorts every image into one of two kinds.

A **real image** forms where light rays *physically converge*. Because the light energy actually arrives at that location, a real image can be caught on a screen, a sheet of film, or the retina of your eye, and it would exist whether or not anyone is there to look. A **virtual image** forms where the rays only *appear* to converge: traced backward, the diverging rays seem to emanate from a point, but no light actually reaches it. A virtual image cannot be projected onto a screen; it lives only in the geometry of the rays as your eye (or a camera) interprets them. The reflection you see in a flat mirror is the most familiar virtual image of all.

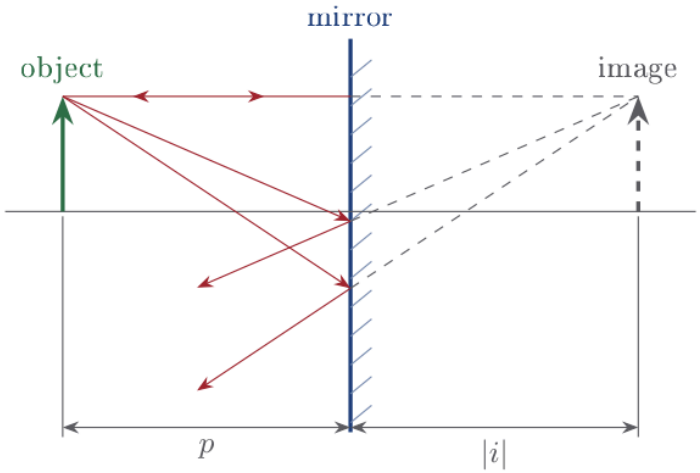


Figure 32.1. Image in a plane mirror. Every ray from the tip of the object reflects as though it came from one point, the same distance behind the mirror as the object is in front. The ray traveling parallel to the axis strikes the mirror head-on and returns along itself; the other two leave the mirror diverging. Traced backward (dashed), all three meet at the tip of the image, which is virtual, upright, and the same size.

•Figure description

A ray diagram for a plane mirror, drawn as a vertical hatched line. A green upright arrow, the object, stands a distance p to the left of it on the horizontal axis. Three red rays leave the tip of the object. One travels parallel to the axis, strikes the mirror head-on, and reflects straight back along itself. The other two strike the mirror at two different points below the axis and reflect down and to the left, spreading apart. Dashed gray lines continue all three reflected rays backwards through the mirror; they meet at the tip of a dashed gray arrow the same height as the object, standing a distance equal to p on the far side. The image is upright, the same size, and virtual.

Consider a point object a distance d in front of a flat, or **plane**, mirror (Figure 32.1). Rays spread out from the object, strike the mirror, and reflect, each obeying the law of reflection from Chapter 31 ($\theta_i = \theta_r$). To your eye the reflected rays diverge, exactly as if they had come from a point *behind* the mirror. A little geometry with the equal angles shows that this apparent source lies the same distance d behind the mirror as the object lies in front:

$$i = -p \qquad \text{(plane mirror)}. \tag{32.1}$$

The negative sign, in the convention we adopt below, flags the image as virtual: it sits behind the mirror, where no light actually goes. The image is upright, the same size as the object, and as far behind the glass as you are in front of it.

Now we can dispose of the old riddle. A plane mirror does *not* swap left for right. What it reverses is *front and back*: the one axis perpendicular to its surface. When you face the mirror, the version of you in it is like a glove turned inside out along the front–back direction: your nose, nearest the glass, stays nearest; the back of your head, farthest, stays farthest, but that single depth flip is what your brain *interprets* as a left–right swap, because you imagine yourself having walked around to stand where your reflection is. Up and down are untouched, and so, in truth, are left and right; only the axis pointing into the mirror is turned around. We will make this precise in the final problems.

Checkpoint 32.1

You want a wall mirror in which you can see your whole body, head to toe. Using the law of reflection and the geometry of Figure 32.1, argue that the mirror need only be *half your height*, and that this minimum size does not depend on how far you stand from it.

Answer

Answers to Checkpoints

Checkpoint 32.1. Your image is as far behind the mirror as you are in front of it. Light from your feet reaches your eye by reflecting off the mirror at the point halfway between the heights of your feet and your eye: the law of reflection makes the two angles equal, so the reflection point is at the midpoint. Likewise, light from the top of your head reflects at the point halfway between your head and your eye. The part of the mirror you need spans half the distance from head to feet: half your height. Moving back does not change this: the midpoints stay at the same heights, so the required mirror is the same size at any distance.

This page titled 32.1: Images and Plane Mirrors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.2: Spherical Mirrors

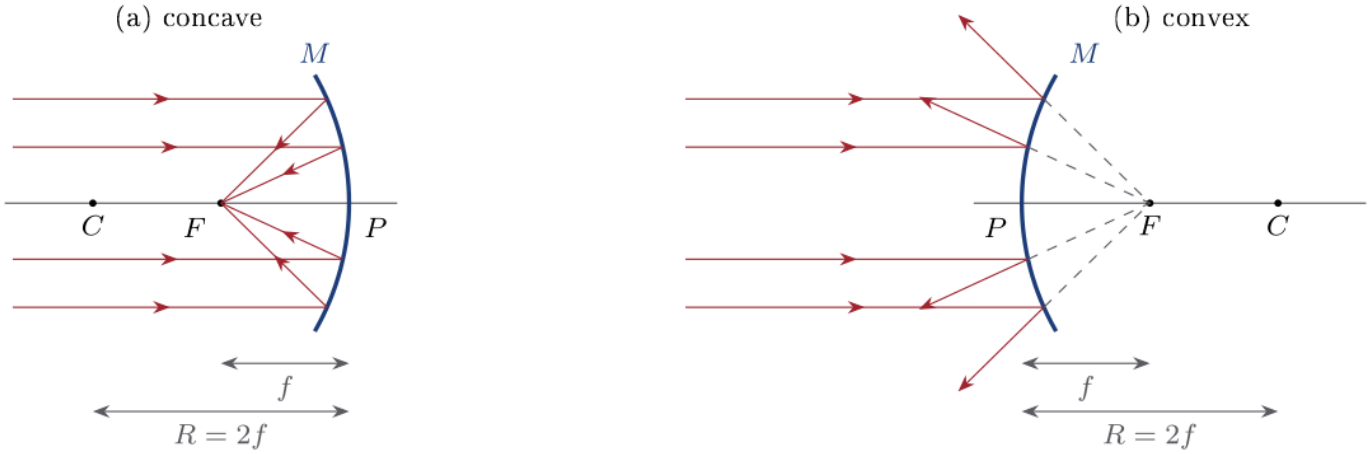


Figure 32.2. Spherical mirrors. (a) A concave mirror reflects parallel rays through a real focal point F in front. (b) A convex mirror reflects them so they diverge, appearing to come from a virtual F behind. In both, F lies halfway to the center of curvature C, so $f = R/2$.

•Figure description

Two panels comparing spherical mirrors. Panel (a), concave, shows a mirror arc curving away from the incoming light, with the center of curvature C and the focal point F marked on the axis in front of it, F halfway between C and the mirror. Four red rays arrive parallel to the axis and, after reflecting, all cross at F. Panel (b), convex, shows an arc curving toward the light, with C and F marked behind the mirror. Parallel rays reflect off it and spread apart; dashed gray lines extend each reflected ray backwards to meet at the virtual F. Both panels mark f from the mirror to F and R equals 2 f from the mirror to C.

Curve a mirror and it gains a power the flat one lacks: it can bring light to a focus and form real images, magnify, or shrink. The simplest curved mirrors are **spherical mirrors**: a small slice cut from the surface of a sphere. Which face you silver decides its type. A **concave** mirror has its reflecting surface on the inside of the curve, cupped toward you like the bowl of a spoon: it is, conveniently, shaped like a *cave*, which is how to remember the name. A **convex** mirror bulges toward you, its reflecting surface on the outside of the curve, like the back of the spoon.

Every spherical mirror has a **center of curvature** C (the center of the sphere it was cut from) and a **focal point** F. The focal point is where rays arriving parallel to the central axis (the **principal axis**) cross after reflecting (Figure 32.2). For a concave mirror these parallel rays genuinely converge through F in front of the mirror; for a convex mirror they spread apart after reflecting, and only their backward extensions meet, at a virtual F behind the mirror. The focal point sits halfway between the mirror and its center of curvature, so the **focal length** is half the radius:

Key Equation: Focal length of a spherical mirror

$$f = \frac{R}{2}. \tag{32.2}$$

Ray diagrams

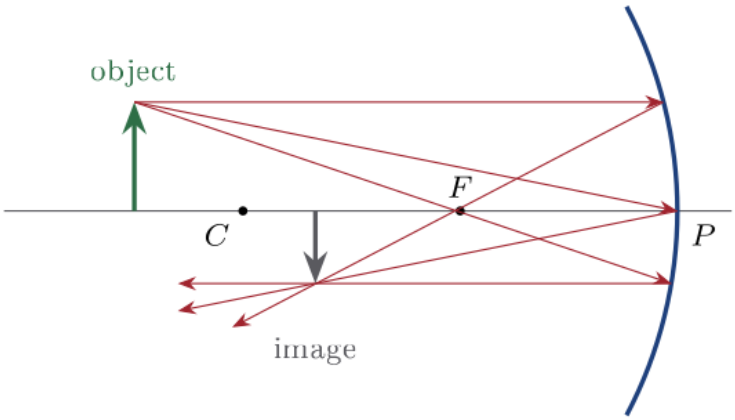


Figure 32.3. Ray diagram for a concave mirror with the object beyond C. The parallel ray reflects through F, the ray through F reflects parallel, and the ray striking the center P reflects at an equal angle on the other side of the axis. All three cross at one point in front of the mirror, forming a *real, inverted, reduced* image.

•Figure description

A ray diagram for a concave mirror with the object beyond the center of curvature. A green upright arrow stands to the left of C on the axis. Three red rays leave its tip. The first travels parallel to the axis, hits the mirror, and reflects back through the focal point F. The second passes through F on the way in, hits the mirror below the axis, and reflects back parallel to the axis. The third strikes the mirror at its center P, where the axis meets it, and reflects at an equal angle below the axis. All three cross at a single point between C and F, the tip of a short gray arrow pointing downward from the axis: the image is real, inverted, and smaller than the object.

To locate an image without algebra, trace two **principal rays** from the top of the object; where they cross (or appear to) is the top of the image. For mirrors:

- A ray arriving *parallel* to the principal axis reflects back *through the focal point* F.
- A ray passing *through the focal point* F reflects back *parallel* to the axis.

Two more rays make useful checks. A ray striking the mirror at its center P reflects at an equal angle on the other side of the axis, just as it would from a plane mirror, and a ray aimed through the center of curvature C strikes the mirror head-on and returns along itself. Any two rays locate the image; a third confirms it. Figure 32.3 traces three rays for an object placed beyond C on a concave mirror; they cross in front of the mirror, so the image is real, inverted, and (here) reduced.

The mirror equation

Ray diagrams show the kind of image; for its exact location and size we use the **mirror equation**, which ties together the focal length f, the object distance p, and the image distance i (all measured from the mirror):

Key Equation: Mirror (and thin-lens) equation

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{i}. \tag{32.3}$$

The size and orientation come from the **magnification**, the ratio of image height h' to object height h:

Key Equation: Magnification

$$m = \frac{h'}{h} = -\frac{i}{p}. \tag{32.4}$$

A positive m means an upright image, negative means inverted; $|m| > 1$ is enlarged, $|m| < 1$ reduced.

Everything hinges on **signs**, the part of this subject that trips up the most students. The intuition is to track the direction the incident light travels: a distance measured on the side the light comes from (and returns to, for a mirror) is positive; a distance on the far, "no-light" side is negative. In practice the convention reduces to the rules in Table 32.1, and the same rules will carry over, unchanged, to lenses.

Table 32.1. Sign convention for mirrors and thin lenses. The same conventions apply to both.

Quantity	Positive (+)	Negative (−)
Object distance p	real object (light side)	virtual object (rare)
Image distance i	real image	virtual image
Focal length f	concave mirror, converging lens	convex mirror, diverging lens
Magnification m	upright image	inverted image

Example 32.1: A book in a convex mirror

A convex mirror has a radius of curvature of $80\ \text{cm}$. A $14\ \text{cm}$ tall book is placed $60\ \text{cm}$ in front of the mirrored side. (a) Where is the image? (b) How tall is it?

Solution

(a) The mirror is convex, so its focal length is negative: $f = R/2 = -80/2 = -40\ \text{cm}$. With the object real, $p = +60\ \text{cm}$. Solve the mirror equation, equation (32.3), for i:

$$\frac{1}{i} = \frac{1}{f} - \frac{1}{p} = \frac{1}{-40} - \frac{1}{60} = -0.0250 - 0.0167 = -0.0417\,\text{cm}^{-1}, \tag{32.2.1}$$

$$i = \boxed{-24\,\text{cm}}. \tag{32.2.2}$$

The negative sign means the image is virtual, sitting $24\ \text{cm}$ *behind* the mirror.

(b) The magnification, equation (32.4), is

$$m = -\frac{i}{p} = -\frac{-24}{60} = +0.40, \tag{32.2.3}$$

so the image is upright (positive m) and reduced. Its height is

$$h' = m\,h = (0.40)(14\,\text{cm}) = \boxed{5.6\,\text{cm}}. \tag{32.2.4}$$

This is exactly why convex mirrors are used as security and passenger-side mirrors: they form small, upright, virtual images and so pack a wide field of view into a little glass, the reason for the etched warning that objects are closer than they appear.

Checkpoint 32.2

A dentist holds a small mirror that makes an enlarged, upright image of a tooth. Must the mirror be concave or convex? Where is the tooth relative to the mirror's focal point? (Reason from Table 32.1 and the sign of m.)

Answer

Answers to Checkpoints

Checkpoint 32.2. Concave. A convex mirror always gives a reduced upright image, so only a concave mirror can magnify. An upright image has $m > 0$, which for a mirror means a virtual image ($i < 0$); a concave mirror forms an enlarged, upright, virtual image only when the object is *inside* the focal point, $p < f$. The dentist holds the mirror closer to the tooth than its focal length.

This page titled 32.2: Spherical Mirrors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.3: Thin Lenses

A **lens** forms images by refraction (Chapter 31) rather than reflection, bending light as it enters and leaves the glass. A **thin lens** is one whose thickness is negligible compared with the distances involved, which lets us treat both bendings as happening at a single plane. As with mirrors there are two families. A **converging** (convex) lens is thicker at its middle and brings parallel rays together at a real focal point on the far side. A **diverging** (concave) lens is thinner at the middle and spreads parallel rays apart, so they appear to come from a virtual focal point on the near side. A lens has a focal point on *each* side, symmetric about its center.

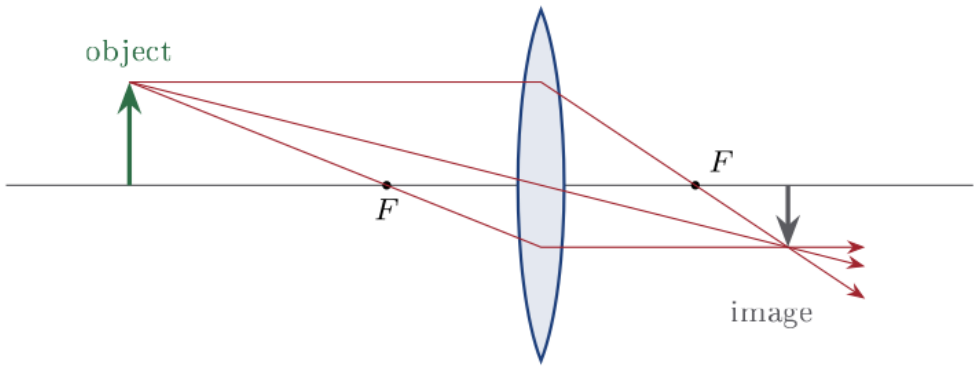


Figure 32.4. Ray diagram for a converging lens, with the object more than twice the focal length away. The parallel ray bends through the far focal point, the central ray passes straight through, and the ray through the near focal point leaves parallel. All three meet on the far side: a *real, inverted, reduced* image.

•Figure description

A ray diagram for a converging lens, drawn as a lens thicker at the middle than at the edges. A focal point F is marked on each side. A green upright object arrow stands on the left, more than twice the focal length from the lens. Three red rays leave its tip. The first travels parallel to the axis and, after the lens, bends down through the far focal point. The second passes straight through the center of the lens undeviated. The third passes through the near focal point on the way in and leaves the lens parallel to the axis. All three meet at one point on the right, between F and twice the focal length, where a gray arrow points downward: a real, inverted, reduced image.

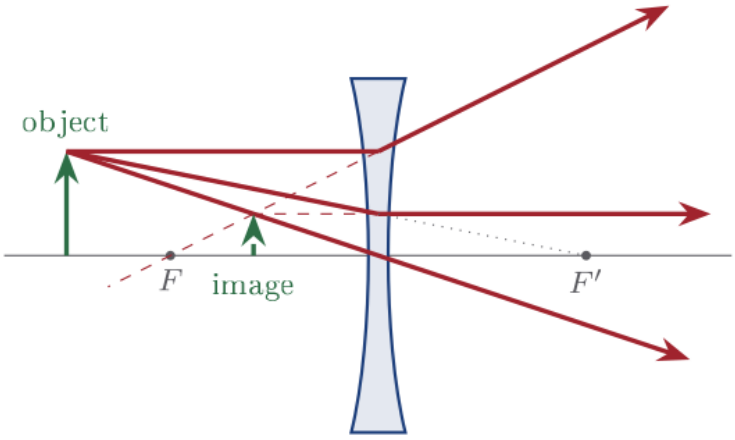


Figure 32.5. Ray diagram for a diverging lens. The central ray passes straight through; the parallel ray bends *away* from the axis as if diverging from the near focal point F; and the ray aimed at the far focal point F' (dotted) leaves parallel. Traced backward (dashed), they meet on the object side: a *virtual, upright, reduced* image.

•Figure description

A ray diagram for a diverging lens, drawn as a lens thinner at the middle than at the edges. A near focal point F is marked on the object side and a far one F prime on the other. A green upright object arrow stands to the left. Three red rays leave its tip. The first passes straight through the center of the lens. The second arrives parallel to the axis and, after the lens, bends away from the axis, upward and to the right; a dashed red line extends it backwards through the near focal point. The third is aimed at the far focal point F prime and leaves the lens parallel to the axis; a dashed red line extends it straight back. The three backward extensions and the central ray meet on the object side at a short dashed green arrow between the lens and the object: a virtual, upright, reduced image.

The wonderful economy of the subject is that a thin lens obeys the *same* equations as a mirror; the thin-lens equation is identical to equation (32.3), and the magnification to equation (32.4), with the very same sign convention of Table 32.1. The only adjustment is geometric: for a lens, light passes *through*, so a real image forms on the side *opposite* the object (whereas a mirror's real image forms on the same side). To trace rays through a lens, use these principal rays:

- A ray through the *center* of the lens passes straight through, undeviated.
- For a *converging* lens: a ray entering parallel to the axis leaves through the far focal point; a ray entering through the near focal point leaves parallel.
- For a *diverging* lens: a ray entering parallel to the axis leaves as if it came from the near focal point; a ray heading toward the far focal point leaves parallel.

Figure 32.4 traces a converging lens with the object beyond twice the focal length; the rays meet on the far side, giving a real, inverted, reduced image, the situation inside a camera. Move the object *inside* the focal point and the same lens behaves quite differently (Figure 32.6): the refracted rays now diverge, and traced backward they meet on the object's own side, giving a virtual, upright, enlarged image. That is how a magnifying glass works.

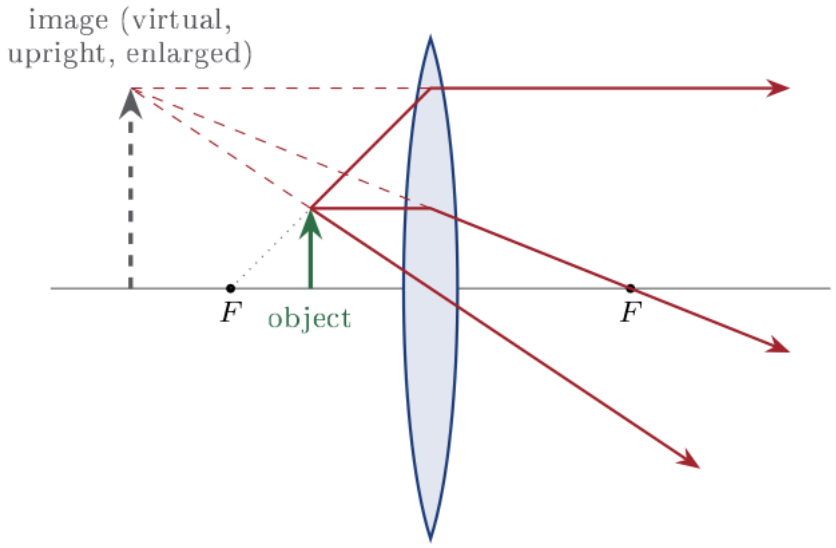


Figure 32.6. A converging lens as a magnifying glass. With the object *inside* the focal point, the three principal rays diverge on the far side. (The third ray travels along the line from the near focal point through the object's tip, dotted, and leaves parallel.) Traced backward (dashed), they meet on the object's side, so the eye sees a *virtual, upright, enlarged* image. Here $p = 0.6f$ gives $i = -1.5f$ and $m = +2.5$.

•Figure description

A ray diagram for a converging lens used as a magnifying glass. A short green upright object stands on the axis between the lens and its near focal point F. Three red rays leave the object's tip. One runs parallel to the axis and bends through the far focal point. One passes straight through the center of the lens. The third travels along the line from the near focal point through the object's tip, reaches the lens high up, and leaves parallel to the axis. On the far side the rays spread apart and never meet. Traced backwards as dashed lines, they meet on the object's side of the lens, farther away than the object and higher up, where a dashed gray upright arrow, about two and a half times as tall as the object, marks the virtual, upright, enlarged image.

A diverging lens tells a different story (Figure 32.5). Its focal length is negative, and it spreads parallel rays apart as though they came from a focal point on the *near* side. Whatever the object distance, a diverging lens always forms a *virtual, upright, reduced* image on the same side as the object, which is why peephole and some eyeglass lenses, used to widen the field or correct nearsightedness, always shrink what you see.

Example 32.2: A converging lens, three placements

A converging lens has a focal length of $10.0\ \text{cm}$. Find and describe the image when an object is placed (a) $30.0\ \text{cm}$, (b) $10.0\ \text{cm}$, and (c) $5.00\ \text{cm}$ from the lens.

Solution

Use the thin-lens equation, $1/i = 1/f - 1/p$, with $f = +10.0\ \text{cm}$ (converging).

(a) $p = 30.0\ \text{cm}$**:**

$$\frac{1}{i} = \frac{1}{10.0} - \frac{1}{30.0} = 0.0667\,\text{cm}^{-1}, \quad i = \boxed{+15.0\,\text{cm}}, \quad m = -\frac{15.0}{30.0} = -0.500. \tag{32.3.1}$$

The image is real ($i > 0$, far side), inverted ($m < 0$), and reduced: the camera case.

(b) $p = 10.0\ \text{cm}$ **(object at** F**):**

$$\frac{1}{i} = \frac{1}{10.0} - \frac{1}{10.0} = 0, \quad i \to \boxed{\infty}. \tag{32.3.2}$$

The refracted rays come out parallel and never meet: *no image forms*. (This is how a spotlight or lighthouse makes a parallel beam: source at the focal point.)

(c) $p = 5.00\ \text{cm}$ **(object inside** F**):**

$$\frac{1}{i} = \frac{1}{10.0} - \frac{1}{5.00} = -0.100\,\text{cm}^{-1}, \quad i = \boxed{-10.0\,\text{cm}}, \quad m = -\frac{-10.0}{5.00} = +2.00. \tag{32.3.3}$$

The image is virtual ($i < 0$, same side as object), upright, and enlarged $2\times$. This is the magnifying glass: bring the object inside the focal length and the lens enlarges it.

The lensmaker's equation

The focal length of a lens is not arbitrary: it is set by the lens's material and the curvature of its two faces. For a thin lens in air of index of refraction n, with surface radii of curvature r_1 and r_2,

Key Equation: Lensmaker's equation

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \tag{32.5}$$

Here r_1 is the radius of the surface the light meets first and r_2 that of the surface it leaves through. Each surface is a piece of a sphere, and its radius is the radius of that sphere (Figure 32.7). Each radius is taken positive if its surface bulges toward the incoming light and negative if it caves inward, so the equation automatically returns a positive f for a converging lens and a negative f for a diverging one.

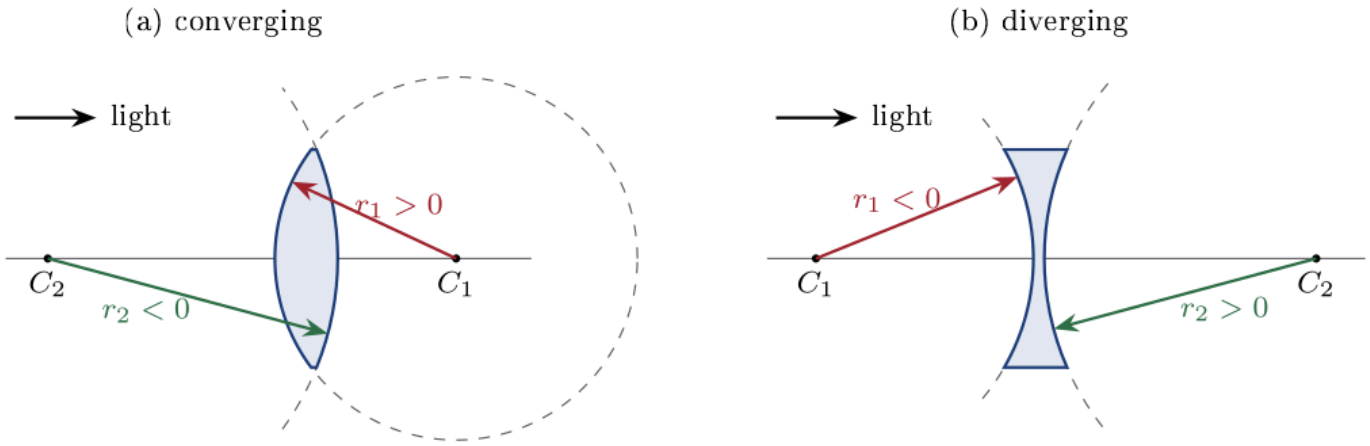


Figure 32.7. Where r_1 and r_2 come from. Each face of a lens is part of a sphere (dashed); its radius is that sphere's radius, measured from the center of curvature C_1 or C_2. (a) A converging lens with unequal faces. The first face bulges toward the incoming light, so $r_1 > 0$; the second curves away from it, so $r_2 < 0$. (b) A diverging lens. Both faces cave in, so the signs reverse: $r_1 < 0$ and $r_2 > 0$.

•Figure description

Two lenses with light arriving from the left, each shown with the two spheres its faces are cut from. Panel a, a converging lens with unequal faces: the left face is strongly curved and the right face gently curved. The left face is part of a small dashed circle whose center C sub 1 lies to the right of the lens; a radius line from C sub 1 to the left face is labeled r sub 1, positive, because that face bulges toward the incoming light. The right face is part of a larger dashed circle whose center C sub 2 lies to the left of the lens; a radius line from C sub 2 to the right face is labeled r sub 2, negative, because that face curves away from the incoming light. Panel b, a diverging lens with unequal faces: the left face caves in, so its center C sub 1 lies to the left and r sub 1 is negative; the right face caves in toward the outgoing side, so its center C sub 2 lies to the right and r sub 2 is positive.

Example 32.3: Grinding a diverging lens

A symmetric biconcave lens is ground from glass of index $n = 1.55$ so that its focal length in air is $20\ \text{cm}$ (in magnitude). Both surfaces have the same radius of curvature; find it.

Solution

A biconcave lens is diverging, so $f = -20\ \text{cm}$. "Symmetric" means both faces share a radius magnitude R; for a biconcave lens the first surface caves in ($r_1 = -R$) and the second, seen from inside the glass, bulges back toward the incoming light ($r_2 = +R$). The lensmaker's equation, equation (32.5), becomes

$$\frac{1}{f} = (n-1)\left(\frac{1}{-R} - \frac{1}{+R}\right) = (n-1)\left(-\frac{2}{R}\right). \tag{32.3.4}$$

Solve for R:

$$R = -2(n-1)\,f = -2(1.55-1)(-20\,\text{cm}) = -2(0.55)(-20\,\text{cm}) = \boxed{22\,\text{cm}}. \tag{32.3.5}$$

Each face is ground to a radius of $22\ \text{cm}$. As a check, a stronger (shorter-f) lens would need a tighter curve (a smaller R), which is just what the equation says.

Checkpoint 32.3

You hold a converging lens at arm's length and see the room behind you small, upside down, and real-looking; you bring the lens close to a printed page and the letters jump up large and right-side up. In terms of the object distance relative to f, explain why the same lens does two such different things. (Compare with parts (a) and (c) of the worked example.)

Answer

Answers to Checkpoints

Checkpoint 32.3. The difference is the object distance. The room is far away, $p > 2f$, so the lens forms a *real*, inverted, reduced image on your side of the lens, which is what you see. The page is close, $p < f$: the rays leaving the lens still diverge, and they appear to come from a *virtual*, upright, enlarged image behind the page. The same lens acts as a camera for distant objects and as a magnifier for objects inside its focal length.

This page titled 32.3: Thin Lenses is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.4: Optical Systems

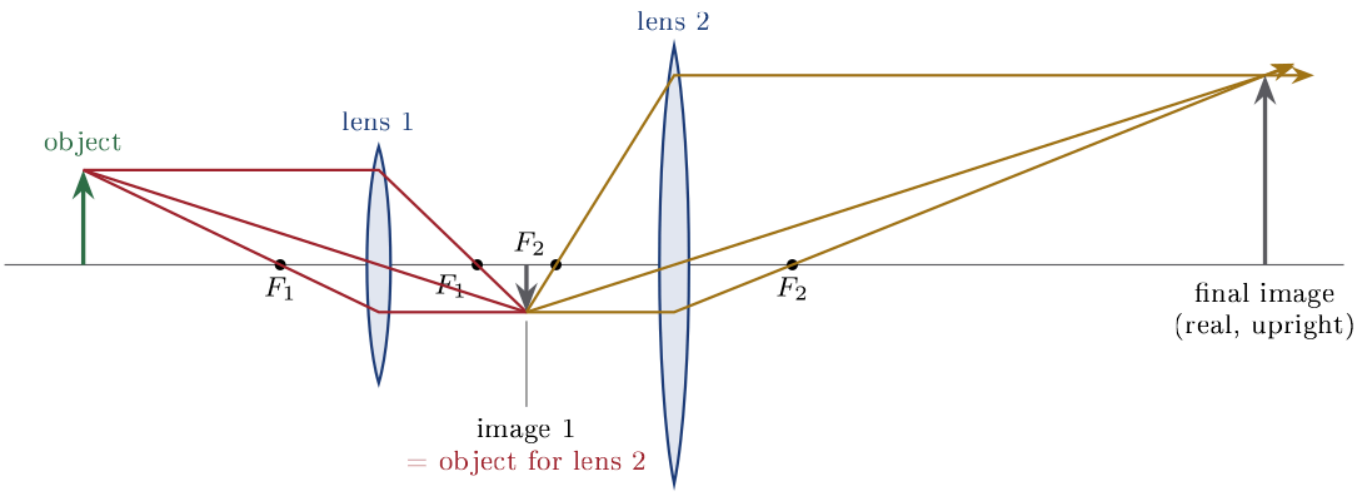


Figure 32.8. An optical system, handled one element at a time. Lens 1 forms an intermediate image (“image 1”); that image then serves as the *object* for lens 2, which forms the final image. Total magnification is the product, $m_{\text{total}} = m_1 m_2$. (Drawn to scale for the worked example below: the final image is real and forms $60\ \text{cm}$ beyond lens 2, far from image 1.)

•Figure description

A to-scale ray diagram of the two-lens example, with light traveling left to right along a horizontal axis. A green upright object stands at the left, then lens 1, then lens 2 further right, with the focal points of each marked. Three red rays from the object's tip pass through lens 1 and meet halfway between the two lenses, forming image 1: a real, inverted, half-size gray arrow, labeled as the object for lens 2. Three gold rays then leave image 1's tip and pass through lens 2: one parallel to the axis that bends through lens 2's far focal point, one straight through its center, and one through its near focal point that leaves parallel. They meet far to the right of lens 2, at the tip of the final image: a real, upright gray arrow twice the height of the object, far from image 1.

Real instruments (microscopes, telescopes, camera zooms, the eye behind your eyeglasses) string several mirrors and lenses together; Section 32.5 looks at the most important ones. These look intimidating but yield to one simple rule: *handle the optics one at a time, in the order the light meets them*. Treat the first optic as though it were alone and find the image it forms. Then use that image as the *object* for the second optic, find the new image, and repeat down the line (Figure 32.8). The work is not harder, only longer.

Two cautions make the bookkeeping reliable. First, the object distance for each optic is measured from *that* optic, so you must account for the spacing between elements. Second, if the first image would land *beyond* the second optic (the light is still converging toward it when it arrives), that image acts as a *virtual object* for the second optic, entering its equation with a negative object distance. The total magnification of the system is the product of the individual magnifications:

$$m_{\text{total}} = m_1 \times m_2 \times \cdots \tag{32.6}$$

Example 32.4: A two-lens system

Two thin converging lenses, of focal lengths $f_1 = 10.0\ \text{cm}$ and $f_2 = 12.0\ \text{cm}$, are separated by $30.0\ \text{cm}$. An object is placed $30.0\ \text{cm}$ to the left of lens 1. Find the position and magnification of the final image.

Solution

Lens 1 alone. With $p_1 = 30.0\ \text{cm}$ and $f_1 = 10.0\ \text{cm}$,

$$\frac{1}{i_1} = \frac{1}{10.0} - \frac{1}{30.0} = 0.0667\,\text{cm}^{-1}, \quad i_1 = +15.0\,\text{cm}, \tag{32.4.1}$$

a real image $15.0\ \text{cm}$ to the *right* of lens 1, with $m_1 = -i_1/p_1 = -15.0/30.0 = -0.500$.

Lens 2. Lens 2 sits $30.0\ \text{cm}$ right of lens 1, so this image lands $30.0 - 15.0 = 15.0\ \text{cm}$ to the *left* of lens 2: a real object for lens 2, $p_2 = +15.0\ \text{cm}$. Then

$$\frac{1}{i_2} = \frac{1}{12.0} - \frac{1}{15.0} = 0.0167\,\text{cm}^{-1}, \quad i_2 = \boxed{+60.0\,\text{cm}}, \tag{32.4.2}$$

a real image $60.0\ \text{cm}$ to the *right* of lens 2, with $m_2 = -i_2/p_2 = -60.0/15.0 = -4.00$.

Combine. The final image lies $60.0\ \text{cm}$ right of lens 2 ($90.0\ \text{cm}$ right of lens 1, and $75.0\ \text{cm}$ beyond image 1), and the total magnification is

$$m_{\text{total}} = m_1 m_2 = (-0.500)(-4.00) = \boxed{+2.00}. \tag{32.4.3}$$

The final image is real, upright, and twice the size of the object. Each lens inverts the image it is handed, so two inversions bring it back upright. Because image 1 sits just outside lens 2's focal length, lens 2 throws the final image far downstream and enlarges it, which is how a projector or a relay lens works. Each step was an ordinary single-lens problem; the system was just two of them in a row.

Checkpoint 32.4

In a two-lens system, the first lens has magnification $m_1 = -2$ and the second $m_2 = -3$. (a) Is the final image upright or inverted, and how large is it compared with the object? (b) If instead the first lens's image would land *beyond* the second lens, what sign does the second lens's object distance p_2 take, and why?

Answer

Answers to Checkpoints

Checkpoint 32.4. (a) The total magnification is $m_1 m_2 = (-2)(-3) = +6$. The image is upright (each lens inverts, and two inversions restore the original orientation) and six times the size of the object. (b) Negative. The light is still converging when it reaches the second lens, so the first image, which it would have formed farther on, lies on the outgoing side of that lens: a *virtual object*. Measured from the second lens, it is on the "wrong" side for a real object, so $p_2 < 0$.

This page titled 32.4: Optical Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.5: Optical Instruments

The rule of Section 32.4 (one optic at a time, each image the object for the next) is all it takes to understand the instruments that extend human sight. We begin with the instrument every one of them is built to serve, the eye, and then look at the magnifying glass, the microscope, and the two kinds of telescope.

The eye

The eye is a camera. Light enters through the clear, curved **cornea**, passes through the pupil, and is focused by the **lens** onto the **retina**, a light-sensitive layer at the back of the eyeball (Figure 32.9a). Most of the bending happens at the cornea, where light passes from air into the watery eye; the lens supplies the adjustable remainder. The image on the retina is real and inverted, and the brain simply interprets it the right way up.

Unlike a camera, the eye cannot move its lens to focus: the distance from lens to retina is fixed at about $2\ \mathrm{cm}$. Instead, small muscles change the *shape* of the lens, and so its focal length. This is **accommodation**. Relaxed, the lens is at its flattest and a normal eye focuses distant objects. Tensed, the lens bulges and focuses nearer ones. The nearest point the eye can bring into focus is its **near point**; for a young adult it is about $25\ \mathrm{cm}$, the standard value used in optics. The farthest is the **far point**, at infinity for a normal eye.

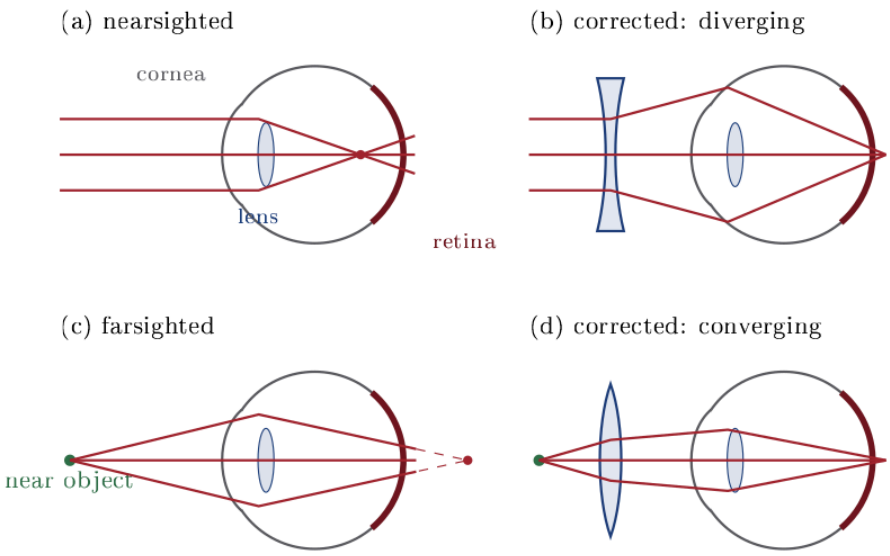


Figure 32.9. Common focusing problems and their correction. (a) A nearsighted eye focuses distant objects in front of the retina, so they look blurred. (b) A diverging lens spreads the incoming rays just enough that the eye focuses them on the retina. (c) A farsighted eye cannot bend the rays from a nearby object strongly enough; they would focus behind the retina. (d) A converging lens supplies the missing bending.

•Figure description

Four panels, each showing an eye in side view with light arriving from the left. The eye is a circle with a bulging cornea at the front, a lens just behind it, and the retina drawn thick along the back. Panel a, nearsighted, uncorrected: three parallel rays from a distant object are focused to a point in front of the retina and have spread apart again by the time they reach it, so the image is blurred. Panel b, nearsighted, corrected: a diverging lens in front of the eye spreads the parallel rays slightly, and the eye now focuses them exactly on the retina. Panel c, farsighted, uncorrected: rays spreading from a nearby point object are aimed at a focus behind the retina, so they are still spread out when they reach it. Panel d, farsighted, corrected: a converging lens in front of the eye makes the rays spread less, and the eye now focuses them on the retina.

When the eye's power does not match its length, vision blurs (Figure 32.9).

- A **nearsighted** (myopic) eye bends light too strongly for its length. Distant objects come to a focus in front of the retina, and the far point is closer than infinity. A *diverging* lens corrects it.
- A **farsighted** (hyperopic) eye bends light too weakly. Its near point is farther than $25\ \mathrm{cm}$, so nearby objects blur. A *converging*

lens corrects it.
- With age the lens stiffens and accommodation weakens, so the near point recedes even in an otherwise normal eye. This is **presbyopia**, the reason most people need reading glasses after about age 45.

The strategy for choosing a corrective lens is always the same: the lens takes an object at the distance the person *wants* to see and forms a virtual image at a distance the person's eye *can* see. The eye then looks at that image instead.

Optometrists describe a lens not by its focal length but by its **power**,

Key Equation: Power of a lens

$$P = \frac{1}{f} \qquad (f \text{ in meters}, P \text{ in diopters}, 1\,\mathrm{D} = 1\,\mathrm{m}^{-1}). \tag{32.7}$$

A converging lens has positive power and a diverging lens negative power, so the sign on a prescription tells you at a glance which problem is being corrected. A strong lens (short focal length) has a large power, and for thin lenses placed close together the powers simply add.

Example 32.5: Prescribing corrective lenses

Treat each corrective lens as sitting right at the eye (a contact lens).

a. A nearsighted student has a far point of $40\ \mathrm{cm}$. What power of lens lets her see distant objects clearly?
b. A farsighted student has a near point of $1.0\ \mathrm{m}$. What power of lens lets him read a book held $25\ \mathrm{cm}$ away?

Solution

(a) The lens must take an object at infinity and form a virtual image at the far point, $i = -0.40\ \mathrm{m}$:

$$P = \frac{1}{f} = \frac{1}{p} + \frac{1}{i} = \frac{1}{\infty} + \frac{1}{-0.40\,\mathrm{m}} = \boxed{-2.5\,\mathrm{D}}. \tag{32.5.1}$$

A diverging lens of focal length $-40\ \mathrm{cm}$, as expected for nearsightedness.

(b) The lens must take an object at $p = 0.25\ \mathrm{m}$ and form a virtual image at the near point, $i = -1.0\ \mathrm{m}$:

$$P = \frac{1}{0.25\,\mathrm{m}} + \frac{1}{-1.0\,\mathrm{m}} = 4.0 - 1.0 = \boxed{+3.0\,\mathrm{D}}. \tag{32.5.2}$$

A converging lens of focal length $33\ \mathrm{cm}$.

Angular magnification and the magnifying glass

How large something *looks* is set not by its actual size but by the angle it spans at your eye, because that angle fixes the size of its image on the retina. A coin at arm's length looks bigger than the Moon. To see detail you bring an object closer, which enlarges the angle, but only as far as your near point. At the near point $N = 25\ \mathrm{cm}$, an object of height h spans the angle

$$\theta \approx \frac{h}{N}. \tag{32.5.3}$$

Now hold a converging lens of focal length f in front of your eye and put the object at its focal point (Figure 32.10). The rays from each point of the object leave the lens parallel, so the image is at infinity and your eye views it fully relaxed. A ray from the top of the object through the center of the lens shows that it now spans the angle $\theta' \approx h/f$. Because a magnifier's job is to enlarge an *angle*, it is rated by its **angular magnification** rather than by the image-size ratio m:

Key Equation: Angular magnification of a simple magnifier

$$M = \frac{\theta'}{\theta} \approx \frac{N}{f} = \frac{25\,\text{cm}}{f} \qquad \text{(image at infinity, relaxed eye).} \tag{32.8}$$

A lens of focal length $5\ \text{cm}$ is a $5\times$ magnifier. (Moving the object slightly inside the focal point, so that the virtual image of Figure 32.6 sits at the near point, raises this to $1 + N/f$, at the cost of a tensed eye.)

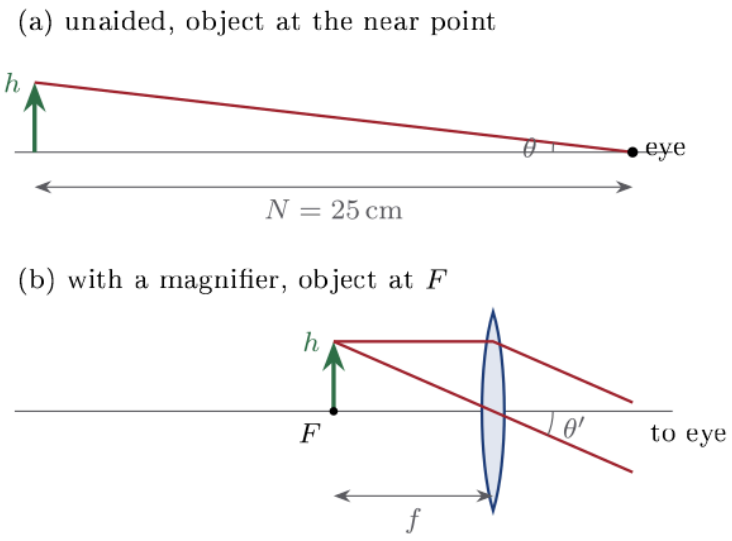


Figure 32.10. Angular magnification. (a) Unaided, the closest an object can be viewed sharply is the near point, where it spans $\theta \approx h/N$. (b) With the object at the focal point of a converging lens held close to the eye, it spans $\theta' \approx h/f$, and the rays leave the lens parallel so the relaxed eye sees it at infinity. The angular magnification is $M = \theta'/\theta = N/f$.

•Figure description

Two panels comparing how large an object looks. Panel a: an eye at the right views a small upright object at the left, a distance N equal to 25 centimeters away. A line from the top of the object to the eye makes a small angle theta with the axis. Panel b: a converging lens sits just in front of the same eye, and the object now stands at the lens's focal point, a distance f away. A ray from the top of the object through the center of the lens makes the larger angle theta prime with the axis, and a second ray leaves the lens parallel to it, toward the eye, showing that the image is at infinity.

The compound microscope

A single magnifier tops out at about $10\times$ to $20\times$, because a much shorter focal length means a tiny, badly distorting lens. A **compound microscope** gets far more by using two lenses in sequence (Figure 32.11). The **objective**, a lens of very short focal length f_o, sits just beyond the specimen and forms a real, inverted, greatly enlarged intermediate image inside the tube. The **eyepiece** is then used as a magnifying glass to view that image, placed so the image falls at its focal point and the eye stays relaxed.

The total magnification is the product of the two stages. The specimen sits just outside the objective's focal point, and the intermediate image forms a distance L, the **tube length**, beyond the objective's far focal point. So $i_o \approx L + f_o$, $p_o \approx f_o$, and the objective's magnification is $m_o = -i_o/p_o \approx -L/f_o$. The eyepiece adds an angular magnification N/f_e, giving

Key Equation: Compound microscope

$$M = m_o \times M_e \approx -\frac{L}{f_o}\,\frac{N}{f_e}, \qquad N = 25\,\text{cm}. \tag{32.9}$$

The negative sign records that the final view is inverted, which is why a slide seems to move the wrong way when you push it. Many microscopes use the standard tube length $L = 160\ \text{mm}$.

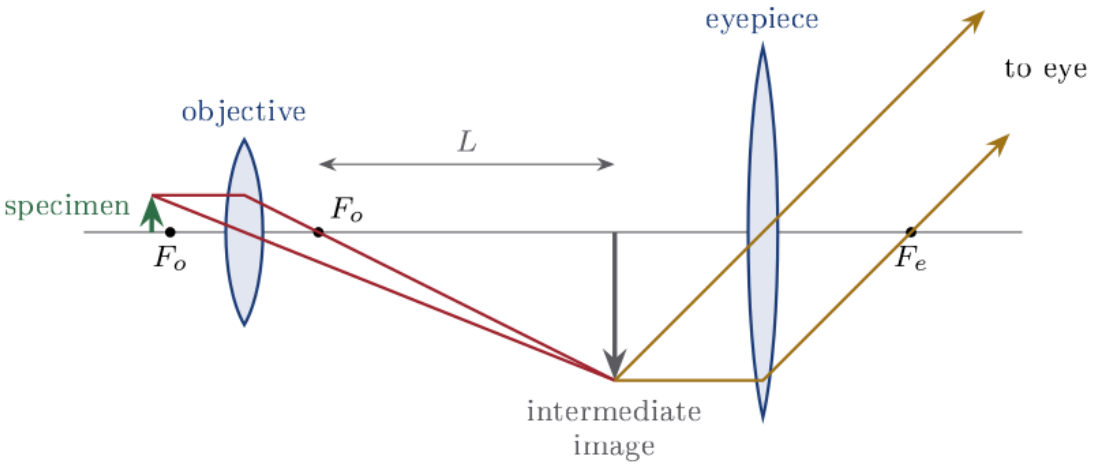


Figure 32.11. A compound microscope (not to scale). The objective forms a real, inverted, enlarged image a tube length L beyond its focal point. That image sits at the focal point of the eyepiece, which acts as a magnifying glass and sends parallel rays to the relaxed eye. The total magnification is $M \approx -(L/f_o)(N/f_e)$.

•Figure description

A ray diagram of a compound microscope, not to scale. A small green specimen stands just to the left of the objective's focal point. Red rays from the top of the specimen pass through the objective, a short-focal-length lens, and meet far to the right at the tip of a large, inverted, real intermediate image. That image sits at the focal point of the eyepiece, a second lens further right. Gold rays from the tip of the intermediate image pass through the eyepiece and leave it parallel to each other, angled upward, toward the eye. The tube length L is marked between the objective's far focal point and the intermediate image.

Example 32.6: A student microscope

A microscope has an objective of focal length $8.0\ \text{mm}$, an eyepiece of focal length $20\ \text{mm}$, and a tube length of $160\ \text{mm}$. (a) Find the overall magnification. (b) How far in front of the objective must the specimen sit?

Solution

(a) The objective magnifies by $m_o \approx -L/f_o = -160/8.0 = -20$, and the eyepiece by $M_e = N/f_e = 250\ \text{mm}/20\ \text{mm} = 12.5$, so

$$M \approx (-20)(12.5) = \boxed{-250}. \tag{32.5.4}$$

The view is enlarged 250 times and inverted.

(b) The intermediate image lies $i_o = L + f_o = 168\ \text{mm}$ beyond the objective. The thin-lens equation gives the object distance:

$$\frac{1}{p_o} = \frac{1}{8.0} - \frac{1}{168} = 0.1190\,\text{mm}^{-1}, \qquad p_o = \boxed{8.4\,\text{mm}}. \tag{32.5.5}$$

The specimen sits just $0.4\ \text{mm}$ outside the focal point, which is why a microscope's focus knob moves in such tiny steps.

Telescopes

A telescope looks at objects so far away that their light arrives as parallel rays. Its **objective** therefore forms a small, real, inverted image at its focal point, and an **eyepiece** placed with its own focal point at the same spot views that image as a magnifier. The two lenses are separated by $f_o + f_e$, and the rays leave the eyepiece parallel, so the eye is relaxed (Figure 32.12).

A distant object spans some angle θ at the telescope. The objective's image of it, at distance f_o, has height $h \approx f_o\theta$; seen through the eyepiece, that image spans $\theta' \approx h/f_e$. The angular magnification is therefore

Key Equation: Angular magnification of a telescope

$$M = \frac{\theta'}{\theta} = -\frac{f_o}{f_e}, \tag{32.10}$$

negative because the view is inverted (a nuisance on Earth, irrelevant for stars). A long objective focal length and a short eyepiece focal length give high magnification, and switching eyepieces changes it.

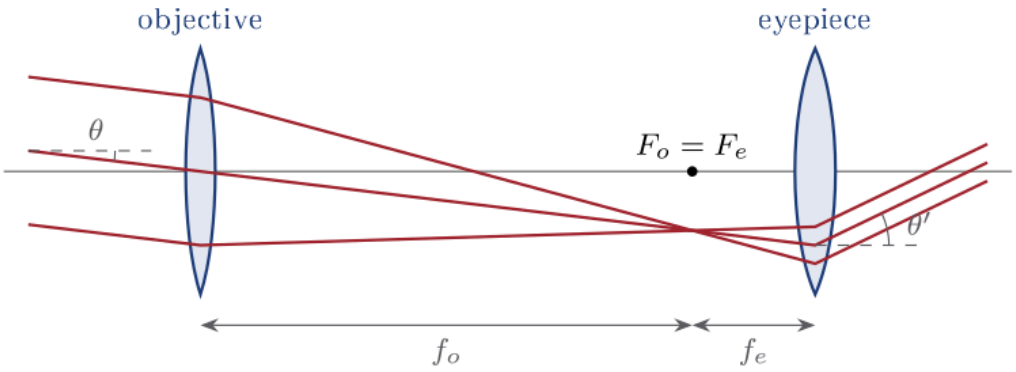


Figure 32.12. A refracting telescope. Parallel light from a distant object, arriving at angle θ, is focused by the objective at its focal point, which is also the eyepiece's focal point. The rays leave the eyepiece parallel again at the larger angle θ'; the angular magnification is $M = -f_o/f_e$ (here -4).

•Figure description

A refracting telescope. Three parallel red rays from a distant object arrive from the left, angled slightly downward at angle theta. They pass through the objective lens and meet at a point a little below the axis, at the shared focal point of the objective and the eyepiece, a distance f sub o from the objective. They continue, spreading, to the eyepiece a distance f sub e further on, and leave it parallel again, now angled upward at the much larger angle theta prime.

Magnification is not actually a telescope's most important job. Most astronomical objects are faint, and the bigger the objective, the more light it collects: light gathered is proportional to the objective's area, and so to the square of its diameter D. A larger objective also blurs fine detail less, because diffraction (Chapter 34) sets a limit on resolution that improves as D grows.

That push for larger objectives is why every major research telescope is a **reflecting telescope**, using a concave mirror instead of a lens as its objective. A lens bends different colors by slightly different amounts (*dispersion*: the index of refraction of glass depends slightly on wavelength), so it focuses each color at a slightly different point; a mirror reflects all colors alike. A lens must also be supported only at its rim and sags under its own weight, while a mirror can be supported across its entire back. The largest refracting telescope ever built for astronomy, at Yerkes Observatory, has a lens just over $1\ \mathrm{m}$ across; the Hubble Space Telescope's mirror is $2.4\ \mathrm{m}$ across, and the James Webb Space Telescope's primary mirror, made of 18 hexagonal segments, is $6.5\ \mathrm{m}$ across.

In Newton's original design (Figure 32.13), the concave **primary mirror** at the back of the tube reflects the incoming light toward its focal point, and a small flat **diagonal mirror** near the front turns the converging beam out through the side of the tube to the eyepiece. The magnification is again $M = -f_o/f_e$, with $f_o = R/2$ the focal length of the primary mirror.

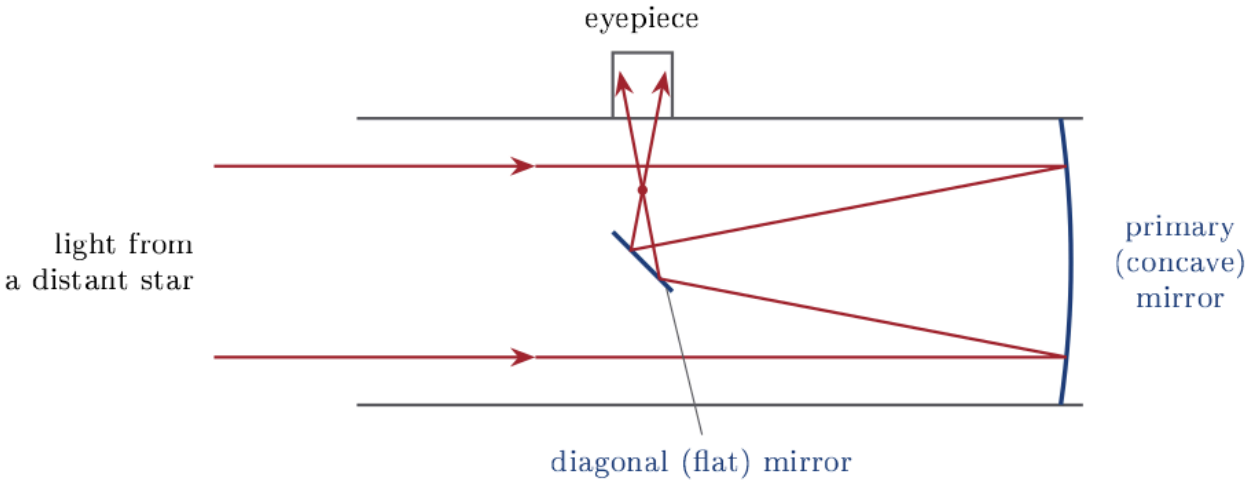


Figure 32.13. A Newtonian reflecting telescope. The concave primary mirror is the objective. A small flat diagonal mirror redirects the converging light to an eyepiece on the side of the tube, where it is viewed exactly as in a refractor.

•Figure description

A Newtonian reflecting telescope, drawn as a long tube open at the left end. Two parallel red rays from a distant star enter the open end, pass on either side of a small flat diagonal mirror near the front, and travel the length of the tube to a concave primary mirror at the back. They reflect from it and converge back up the tube toward a focus, but before reaching it they strike the diagonal mirror, set at 45 degrees, which turns them upward to a focus just inside an eyepiece mounted on the side of the tube.

Example 32.7: A backyard reflector

An amateur's Newtonian telescope has a primary mirror $20.0\ \text{cm}$ in diameter with a radius of curvature of $2.40\ \text{m}$. It comes with eyepieces of focal length $25\ \text{mm}$ and $10\ \text{mm}$. (a) Find the magnification with each eyepiece. (b) How many times more light does the mirror collect than a dark-adapted eye, whose pupil is about $7.0\ \text{mm}$ across?

Solution

(a) The primary mirror's focal length is $f_o = R/2 = 1.20\ \text{m} = 1200\ \text{mm}$. Then

$$M_{25} = -\frac{1200\,\text{mm}}{25\,\text{mm}} = \boxed{-48}, \qquad M_{10} = -\frac{1200\,\text{mm}}{10\,\text{mm}} = \boxed{-120}. \tag{32.5.6}$$

Swapping eyepieces is how an observer changes magnification.

(b) Collected light is proportional to aperture area, so to the square of the diameter:

$$\frac{D^2_{\text{mirror}}}{D^2_{\text{pupil}}} = \left(\frac{200\,\text{mm}}{7.0\,\text{mm}}\right)^2 = \boxed{816}. \tag{32.5.7}$$

The telescope shows stars about 800 times fainter than the eye alone can see, whichever eyepiece is in place. That, more than magnification, is what a telescope is for.

Checkpoint 32.5

(a) An eyeglass prescription reads $-3.0\ \text{D}$. Is the wearer nearsighted or farsighted, and what is the lens's focal length? (b) You swap a telescope's $25\ \text{mm}$ eyepiece for a $10\ \text{mm}$ one. What happens to the magnification? Does the telescope now gather more light?

Answer

Answers to Checkpoints

Checkpoint 32.5. (a) Negative power means a diverging lens, which corrects nearsightedness. Its focal length is $f = 1/P = 1/(-3.0\ \mathrm{D}) = -0.33\ \mathrm{m}$, or $-33\ \mathrm{cm}$. (b) Since $M = -f_o/f_e$, cutting f_e from 25 to $10\ \mathrm{mm}$ raises the magnification by a factor of 2.5. The light gathered does not change: it is set by the area of the objective, not by the eyepiece. The image is larger but no brighter, and in fact the same light is spread over a larger apparent area.

This page titled 32.5: Optical Instruments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.S: Images (Summary)

Summary

- An **image** forms where rays from each object point reconverge. A **real image** (rays physically meet) can be projected on a screen; a **virtual image** (rays only appear to meet) cannot.
- A **plane mirror** gives a virtual, upright, same-size image as far behind as the object is in front ($i = -p$). It reverses front–back, not left–right.
- **Spherical mirrors** have $f = R/2$ ($+$ concave, $-$ convex). Locate images with the **mirror equation** $1/f = 1/p + 1/i$ and **magnification** $m = h'/h = -i/p$, using the sign convention of Table 32.1.
- **Thin lenses** obey the *same* two equations and sign convention; converging lenses have $f > 0$, diverging $f < 0$. A lens's real image forms on the side opposite the object. The **lensmaker's equation** $1/f = (n-1)(1/r_1 - 1/r_2)$ sets f from the glass and curvatures.
- Trace images with **principal rays**; $m > 0$ upright, $m < 0$ inverted, $|m| > 1$ enlarged.
- Analyze an **optical system** one element at a time (each image is the next optic's object) and multiply magnifications, $m_{\text{total}} = m_1 m_2 \cdots$
- The **eye** focuses by changing the shape of its lens; its near point is about $25\ \text{cm}$. Nearsightedness is corrected with a diverging lens and farsightedness with a converging one. A lens's **power** is $P = 1/f$ in diopters.
- Optical instruments are rated by **angular magnification** $M = \theta'/\theta$: a magnifier gives $M \approx N/f$, a compound microscope $M \approx -(L/f_o)(N/f_e)$, and a telescope $M = -f_o/f_e$. Large telescopes use mirrors, which gather more light without color errors.

Key Equations

Plane mirror

$$i = -p \qquad \text{(virtual, upright, same size)} \tag{32.S.1}$$

Focal length of a spherical mirror

$$f = \frac{R}{2} \qquad (+\ \text{concave},\ -\ \text{convex}) \tag{32.S.2}$$

Mirror and thin-lens equation

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{i} \tag{32.S.3}$$

Magnification

$$m = \frac{h'}{h} = -\frac{i}{p} \qquad \begin{array}{l} m > 0 \text{ upright},\ m < 0 \text{ inverted} \\ |m| > 1 \text{ enlarged},\ |m| < 1 \text{ reduced} \end{array} \tag{32.S.4}$$

Lensmaker's equation

$$\frac{1}{f} = (n-1)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{32.S.5}$$

Optical system: total magnification

$$m_{\text{total}} = m_1 \times m_2 \times \cdots \tag{32.S.6}$$

Power of a lens

$$P = \frac{1}{f} \qquad (\text{diopters},\ f \text{ in m}) \tag{32.S.7}$$

Simple magnifier (relaxed eye)

$$M = \frac{\theta'}{\theta} \approx \frac{N}{f}, \qquad N = 25\,\text{cm} \tag{32.S.8}$$

Compound microscope

$$M \approx -\frac{L}{f_o}\,\frac{N}{f_e} \tag{32.S.9}$$

Telescope

$$M = -\frac{f_o}{f_e} \tag{32.S.10}$$

This page titled 32.S: Images (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

32.E: Images (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Unless stated otherwise, use the sign convention of Table 32.1: real objects and images give positive p, i; converging optics have $f > 0$, diverging $f < 0$; $m = -i/p$.

Section 32.1 — Images and Plane Mirrors

32.1 ★ *[Conceptual.]* State the difference between a real and a virtual image, and give one everyday example of each.

32.2 ★ You stand $1.2\ \text{m}$ in front of a plane mirror. (a) How far behind the mirror is your image? (b) How far apart are you and your image?

32.3 ★★ *[Conceptual.]* A small mirror is mounted on a wall. Using ray geometry, show that the part of your *own body* you can see in it stays the same as you walk toward or away from the mirror. Then explain why the view of the room behind you does change.

32.4 ★★ Two plane mirrors hang on opposite walls of a room, parallel to each other and $4.0\ \text{m}$ apart. You stand $1.5\ \text{m}$ from the left mirror and look into it, where you see a long row of images of yourself.

a. How far from you is the nearest image?
b. The next image is the left mirror's image of your image in the right mirror. How far from you is it?
c. How far from you is the third image in the row?

Section 32.2 — Spherical Mirrors

32.5 ★ A concave mirror has a radius of curvature of $24\ \text{cm}$. What is its focal length?

32.6 ★★ An object is placed $30\ \text{cm}$ in front of a concave mirror of focal length $10\ \text{cm}$. Find (a) the image distance and (b) the magnification, and (c) describe the image (real/virtual, upright/inverted, enlarged/reduced).

32.7 ★★ A $5.0\ \text{cm}$ tall object stands $15\ \text{cm}$ in front of a convex mirror of focal length $20\ \text{cm}$ (in magnitude). Find the image distance, the magnification, and the image height.

32.8 ★★★ *[Real-world application: automotive.]* A convex passenger-side mirror has a focal length of $60\ \text{cm}$ (magnitude). A car is $12\ \text{m}$ behind it. (a) Find the image distance and magnification. (b) The mirror is etched "objects in mirror are closer than they appear." Explain, using your magnification, why the car looks farther away than it is.

32.9 ★★ A concave mirror has a radius of curvature of $40\ \text{cm}$, and an object sits $60\ \text{cm}$ in front of it.

a. Find the focal length.
b. Find the image distance and the magnification.
c. Is the image real or virtual, upright or inverted?

32.10 ★★ A convex (diverging) mirror has a focal length of $-15\ \text{cm}$. An object is placed $30\ \text{cm}$ in front of it. Find the image distance and magnification, and describe the image.

Section 32.3 — Thin Lenses

32.11 ★ A converging lens has a focal length of $15\ \text{cm}$. An object is placed $45\ \text{cm}$ away. Find the image distance.

32.12 ★★ A diverging lens has a focal length of $10.0\ \text{cm}$ (magnitude). An object is placed $30.0\ \text{cm}$ away. Find the image distance and magnification, and describe the image.

32.13 ★★ *[Real-world application: biology.]* A magnifying glass is a converging lens of focal length $8.0\ \text{cm}$. To view a specimen

enlarged and upright, it is placed $5.0\ \text{cm}$ from the lens. Find the image distance and magnification, and state whether the image is real or virtual.

32.14 ★★★ A symmetric biconvex lens is ground from glass of index 1.50. Each face has a radius of curvature of magnitude $30\ \text{cm}$. Find the focal length of the lens. *(Hint: for a biconvex lens, $r_1 = +R$ and $r_2 = -R$.)*

32.15 ★★ A converging lens has a focal length of $12\ \text{cm}$, and an object is placed $18\ \text{cm}$ from it.

a. Where is the image?
b. What is the magnification?
c. Is the image real or virtual?

32.16 ★★ An object is placed $8.0\ \text{cm}$ from a converging lens of focal length $12\ \text{cm}$ (inside the focal point). (a) Find the image distance and magnification. (b) This is how a magnifying glass works: describe the image.

32.17 ★★★ A diverging lens has a focal length of $-20\ \text{cm}$. (a) For an object $30\ \text{cm}$ away, find the image distance and magnification. (b) Show that a diverging lens forms a virtual, upright, reduced image of a real object for *any* object distance.

Section 32.4 — Optical Systems & the Mirror Puzzle

32.18 ★★ Two converging lenses, each of focal length $12\ \text{cm}$, are separated by $40\ \text{cm}$. An object sits $18\ \text{cm}$ to the left of the first lens. Find the position of the final image relative to the second lens.

32.19 ★★★ *[Real-world application: optics.]* A simple telescope uses an objective lens of focal length $100\ \text{cm}$ and an eyepiece of focal length $5.0\ \text{cm}$. Distant starlight enters the objective effectively parallel (object at infinity). (a) Where does the objective form its image? (b) If the eyepiece is placed so this image falls at the eyepiece's focal point, describe the rays leaving the eyepiece and explain why this is comfortable for a relaxed eye.

32.20 ★★★ *[Conceptual — the mirror puzzle.]* Return to the chapter opener. A plane mirror seems to swap your left and right but not your top and bottom.

a. Hold this page up to a mirror. Are the letters of ambulance reversed left–right, or is each letter flipped about a vertical line as if seen from behind?
b. Explain, using the idea that a plane mirror reverses only the axis perpendicular to its surface, why the apparent "left–right" reversal is really a front–back reversal that your brain reinterprets.
c. A mirror lying flat on the floor, which you look down into, reverses what your brain reads as up–down instead. Reconcile this with part (b).

Section 32.5 — Optical Instruments

32.21 ★ (a) An eyeglass lens has a power of $+2.5\ \text{D}$. What is its focal length, and is it converging or diverging? (b) What is the power of a lens with a focal length of $-50\ \text{cm}$?

32.22 ★★ *[Real-world application: vision — accommodation.]* Model the eye as a single thin lens a fixed $2.0\ \text{cm}$ in front of the retina.

a. What power must the eye have to focus a very distant object on the retina?
b. What power must it have to focus an object at the near point, $25\ \text{cm}$ away?
c. By how many diopters must the eye's power change between the two? (This is the eye's range of accommodation.)

32.23 ★★ A nearsighted student has a far point of $80\ \text{cm}$ and a near point of $15\ \text{cm}$.

a. What power of contact lens lets her see distant objects clearly?
b. While she wears these lenses, what is the closest distance at which she can see an object clearly?

32.24 ★★ *[Real-world application: a jeweler's loupe.]* A jeweler examines a gem $3.0\ \text{mm}$ across with a loupe, a converging lens of focal length $2.5\ \text{cm}$, held close to the eye.

a. What angle does the gem span when viewed without the loupe at the near point, $25\ \text{cm}$ away?
b. What angle does it span through the loupe, with the gem at the focal point?
c. Find the angular magnification with a relaxed eye, and the larger value obtained when the image is placed at the near point instead.

32.25 ★★ A refracting telescope has an objective of focal length $90\ \text{cm}$ and an eyepiece of focal length $2.0\ \text{cm}$, adjusted for a relaxed eye.

a. Find its angular magnification.
b. How far apart are the two lenses?
c. The full Moon spans about 0.52° in the sky. What angle does it appear to span through the telescope?

32.26 ★★ A compound microscope has an objective of focal length $6.0\ \text{mm}$, an eyepiece of focal length $20\ \text{mm}$, and a tube length of $160\ \text{mm}$.

a. Find the magnification of the objective, the angular magnification of the eyepiece, and the overall magnification.
b. Keeping the same objective, what eyepiece focal length would give an overall magnification of -500?

32.27 ★★★ *[Real-world application: amateur astronomy.]* A Newtonian reflector has a primary mirror $25\ \text{cm}$ in diameter with a radius of curvature of $2.5\ \text{m}$.

a. What is the focal length of the primary mirror?
b. Find the magnification with an eyepiece of focal length $10\ \text{mm}$.
c. What eyepiece focal length would give a magnification of -50, better for viewing a large faint galaxy?
d. A dark-adapted pupil is about $7.0\ \text{mm}$ across. How many times more light does the mirror collect? A star is 400 times too faint to see with the naked eye: can it be seen through this telescope?

Challenge Problems

32.28 ★★★ *[Real-world application: vision — prescribing eyeglasses.]* Optometrists specify a lens by its *power* $P = 1/f$, measured in diopters ($1\ \text{D} = 1\ \text{m}^{-1}$) when f is in meters. A corrective lens worn close to the eye must take an object at the distance the patient *wants* to see and form an image at the distance the patient's unaided eye *can* see.

a. A patient with presbyopia has a near point of $75\ \text{cm}$ but wants to read at $25\ \text{cm}$. The lens must take an object at $p = 25\ \text{cm}$ and form a virtual image at the near point, $i = -75\ \text{cm}$. Find the required focal length and power, and state whether the lens is converging or diverging.
b. A second patient is nearsighted, with a far point of $2.0\ \text{m}$: objects beyond that are blurred. The lens must take an object at infinity and form a virtual image at the far point. Find its focal length and power.
c. The second patient, now older, also develops a near point of $60\ \text{cm}$. Find the power needed for reading at $25\ \text{cm}$, and explain why a single lens cannot serve both needs, and what a bifocal does about it.
d. An optician grinds the first patient's lens from glass of index 1.52 with one flat face. Using the lensmaker's equation, find the radius of curvature of the curved face.

32.29 ★★★ *[Real-world application: photography.]* A camera uses a thin lens of focal length $50.0\ \text{mm}$ and a sensor $36\ \text{mm} \times 24\ \text{mm}$. A person $1.80\ \text{m}$ tall stands $3.00\ \text{m}$ from the lens.

a. How far behind the lens must the sensor sit for a sharp image?
b. Find the magnification and the height of the image. Will the person fit within the $36\ \text{mm}$ dimension of the sensor? Within the $24\ \text{mm}$ dimension?
c. How far away must the person stand for the image to be exactly $24\ \text{mm}$ tall: a full-length portrait in landscape orientation?
d. Find the sensor distance for a subject at infinity, and compare it with your answer to (a). Use the comparison to explain why a lens barrel needs only a few millimeters of travel to focus from infinity down to a few meters, but a great deal more travel for close-up (macro) work.

32.30 ★★★ *[Real-world application: microscopy.]* A compound microscope has an objective of focal length $f_o = 4.00\ \text{mm}$ and an eyepiece of focal length $f_e = 25.0\ \text{mm}$. A specimen is placed $4.10\ \text{mm}$ in front of the objective.

a. Find the position of the intermediate image formed by the objective, and the objective's magnification. State whether the

intermediate image is real or virtual, upright or inverted.

b. The eyepiece is positioned so that the intermediate image falls at its focal point, sending parallel rays to the relaxed eye. Its angular magnification is then $M_e = (25\ \text{cm})/f_e$. Find M_e, and the total magnification of the instrument.
c. How far apart are the objective and the eyepiece?
d. Suppose you try to reach $2000\times$ by using an objective of focal length $0.80\ \text{mm}$ with the same eyepiece and tube length. The image will indeed be $2000\times$ larger, and yet no new detail will appear. Explain why, referring forward to the diffraction limit of Chapter 34.

This page titled 32.E: Images (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33: INTERFERENCE

In 1801 Thomas Young set out to settle an argument that had run for more than a century. Newton had insisted that light was a stream of tiny particles (corpuscles) that flew in straight lines. A few dissenters, Huygens among them, held that light was a wave. The question seemed almost impossible to decide: light shows none of the obvious wave behavior of water or sound. Then Young sent light through two narrow, closely spaced slits and let it fall on a screen. If light were a hail of particles, he should have seen two bright bands, one behind each slit: the pattern that paint-covered balls thrown at a wall would make. Instead he saw a whole row of bright and dark stripes, many more than two, fading off to either side.

There is only one way to make stripes like that: the light from the two slits must be *adding* in some places and *canceling* in others, exactly as two sets of ripples do where they cross on a pond. Light is a wave. This chapter is about what follows from that single fact: how overlapping light waves reinforce and cancel, a phenomenon called **interference**, and how we read the resulting patterns to measure wavelengths, film thicknesses, and distances to a fraction of a wavelength. We will end with the instrument, the interferometer, that turns interference into the most precise measuring tool ever built: precise enough, two centuries after Young, to detect a passing gravitational wave that changes a $4\ \mathrm{km}$ arm by a small fraction of the width of a proton. And we will circle back to the deepest puzzle of all: the same experiment, run with electrons, makes the same stripes.

Learning Objectives

- Explain the conditions for coherent light to interfere constructively or destructively.
- Locate the bright and dark fringes of Young's double-slit experiment.
- Find the intensity at any point in the double-slit pattern using phasor addition.
- Analyze interference in thin films, accounting for reflection phase shifts, including soap films and anti-reflection coatings.
- Relate a mirror's displacement in a Michelson interferometer to a countable number of fringes.

We will return to Young's experiment, and its unsettling sequel with electrons, in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

33.1: Young's Double-Slit Experiment
33.2: Interference and Double-Slit Intensity
33.3: Interference from Thin Films
33.4: Interferometry
33.S: Interference (Summary)
33.E: Interference (Problems)

This page titled 33: Interference is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.1: Young's Double-Slit Experiment

Chapter 31 described how a light wave moves forward by Huygens' principle, and why its wave nature stays hidden whenever it passes through openings much larger than its wavelength. To reveal it, light must pass through openings comparable to its wavelength, where it spreads out and overlaps with itself. Even then, overlapping light does not automatically produce a visible pattern. For two beams to produce a *stable* interference pattern they must be **coherent**: they must keep a fixed phase relationship over time. Two separate light bulbs never interfere visibly, because each emits in random, ever-shifting bursts; their pattern flickers away faster than any eye or detector can catch. Young's trick was to illuminate *both* slits with light from a *single* source, so that whatever the phase does, it does identically at the two slits. A laser does this for us today.

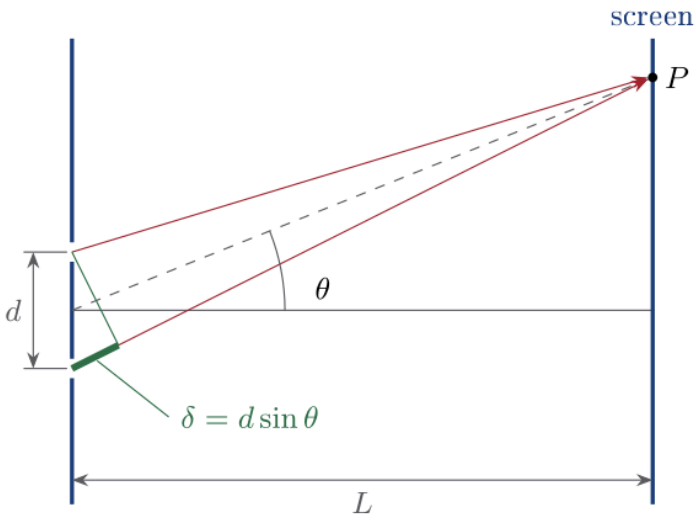


Figure 33.1. Young's double slit. Light from the two slits (separation d) reaches a distant point P at angle θ. The wave from the lower slit travels an extra path $\delta = d\sin\theta$; whether P is bright or dark depends on how this compares with λ.

•Figure description

Young's double-slit geometry. Two slits, one above the other and separated by a distance d, are at the left; a screen stands a distance L away at the right. Two red rays, one from each slit, run up and to the right at the same small angle theta above the central axis, meeting the screen at a point P. Because the rays are nearly parallel, a small green right triangle at the slits shows that the ray from the lower slit must travel an extra distance d sine theta before it sets off -- the path difference that decides whether P is bright or dark.

In Young's experiment, coherent light of wavelength λ falls on two narrow slits separated by a distance d, and the overlapping light lands on a screen a distance L away (Figure 33.1). Consider the light reaching a point P on the screen at angle θ from the central axis. Because the two slits are not equidistant from P, the wave from the farther slit travels an extra distance: the **path difference**. When the screen is far away ($L \gg d$), the two rays to P are very nearly parallel, and a little geometry gives the path difference as

$$\delta = d\sin\theta. \tag{33.1}$$

If this extra path is a whole number of wavelengths, the two waves arrive crest-to-crest and reinforce: a **bright fringe** (constructive interference). If it is a half-integer number of wavelengths, they arrive crest-to-trough and cancel: a **dark fringe** (destructive interference):

Key Equation: Double-slit fringe conditions

$$d\sin\theta = m\lambda, \qquad (\text{bright, maxima}) \tag{33.2}$$
$$d\sin\theta = \left(m + \tfrac{1}{2}\right)\lambda, \qquad (\text{dark, minima}) \tag{33.3}$$

with the order $m = 0, 1, 2, \ldots$

The central bright fringe ($m = 0$) sits straight ahead, where the paths are equal. For points near the center the angle is small, so $\sin\theta \approx \tan\theta = y/L$, where y is the distance up the screen, measured from the center of the central bright fringe (Figure 33.2). The bright fringes then fall at

$$y_{\text{bright}} = m\,\frac{\lambda L}{d}, \tag{33.4}$$

evenly spaced, with adjacent bright fringes separated by

$$\Delta y = \frac{\lambda L}{d}. \tag{33.5}$$

The dark fringes fall halfway between the bright ones, at $y_{\text{dark}} = \left(m + \frac{1}{2}\right)\lambda L/d$. Figure 33.2 shows both sets of fringes on the screen: y_m locates a single fringe measured from the center, while Δy is the gap between any two neighbors. The pattern is a direct ruler for the wavelength: measure the fringe spacing, the slit separation, and the screen distance, and you have λ.

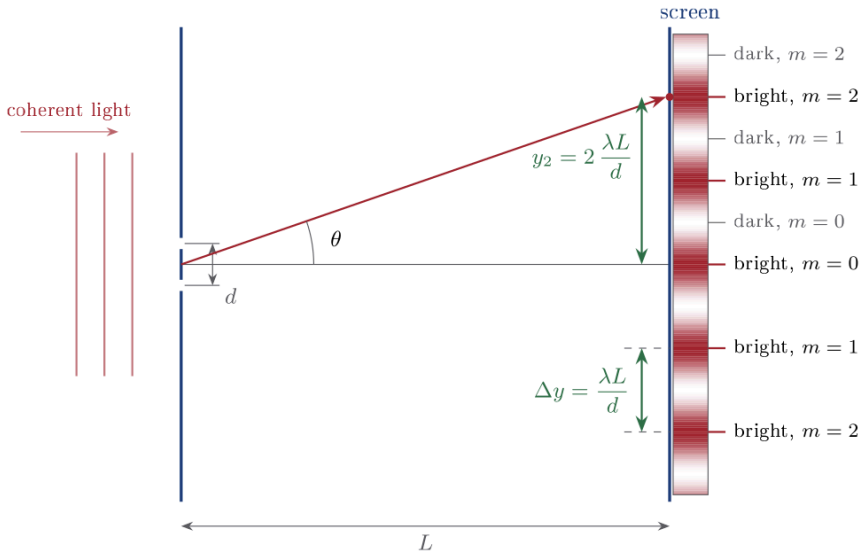


Figure 33.2. The fringes on the screen. Bright fringes ($d\sin\theta = m\lambda$) alternate with dark fringes ($d\sin\theta = (m + \frac{1}{2})\lambda$), with the $m = 0$ bright fringe on the axis. The distance y_m from the center locates a single fringe, here the $m = 2$ bright fringe at $y_2 = 2\lambda L/d$; the spacing $\Delta y = \lambda L/d$ separates any two adjacent bright fringes (or adjacent dark ones). The vertical scale is greatly exaggerated: in practice L is thousands of times larger than d.

•Figure description

The fringe pattern on the screen of a double slit, with distances labeled. Coherent light, drawn as parallel wavefronts, arrives from the left at a barrier with two slits a distance d apart. A screen stands a distance L to the right. On the screen a vertical strip shows the pattern: red bands for bright fringes alternating with white bands for dark fringes, evenly spaced. Labels beside the strip mark, from the center upward, bright m equals 0 on the axis, dark m equals 0, bright m equals 1, dark m equals 1, bright m equals 2 and dark m equals 2; below the axis the bright fringes m equals 1 and 2 are marked too. A red line runs from the slits at angle theta to the m equals 2 bright fringe, and a green double arrow beside the screen measures its height above the axis, labeled y sub 2 equals 2 lambda L over d. Lower down, a second green double arrow spans two adjacent bright fringes, labeled delta y equals lambda L over d.

Example 33.1: Measuring a wavelength

A screen sits $L = 4.80\ \text{m}$ from a double slit whose slits are $d = 0.0300\ \text{mm}$ apart. The first dark fringe appears $4.50\ \text{cm}$ from the center line. (a) Find the wavelength. (b) Find the spacing between adjacent bright fringes.

Solution

(a) The first dark fringe is the $m = 0$ minimum, at $d\sin\theta = \frac{1}{2}\lambda$. Using $\sin\theta \approx y/L$,

$$\frac{1}{2}\lambda = d\,\frac{y}{L} \;\Rightarrow\; \lambda = \frac{2dy}{L} = \frac{2(3.00\times10^{-5}\,\text{m})(0.0450\,\text{m})}{4.80\,\text{m}} = 5.63\times10^{-7}\,\text{m} = \boxed{563\,\text{nm}}, \tag{33.1.1}$$

a yellow-green light.

(b) The bright-fringe spacing is

$$\Delta y = \frac{\lambda L}{d} = \frac{(5.625\times10^{-7}\,\mathrm{m})(4.80\,\mathrm{m})}{3.00\times10^{-5}\,\mathrm{m}} = \boxed{9.00\times10^{-2}\,\mathrm{m} = 9.00\,\mathrm{cm}}. \tag{33.1.2}$$

As a check, the first dark fringe should sit halfway between the central and first bright fringes, at $\frac{1}{2}(9.00\ \mathrm{cm}) = 4.50\ \mathrm{cm}$, exactly as given.

Example 33.2: Two colors, two patterns

A source emits two wavelengths, $\lambda = 430\ \mathrm{nm}$ and $\lambda' = 510\ \mathrm{nm}$, into a double slit with $L = 1.50\ \mathrm{m}$ and $d = 0.0250\ \mathrm{mm}$. How far apart are the two third-order ($m = 3$) bright fringes?

Solution

Each color makes its own pattern; using $y_{\mathrm{bright}} = m\lambda L/d$ for $m = 3$,

$$y_3 = \frac{3\lambda L}{d}, \qquad y_3' = \frac{3\lambda' L}{d}. \tag{33.1.3}$$

Their separation is

$$\Delta y = \frac{3L}{d}(\lambda' - \lambda) = \frac{3(1.50\,\mathrm{m})}{2.50\times10^{-5}\,\mathrm{m}}\,(510-430)\times10^{-9}\,\mathrm{m} = \boxed{1.44\times10^{-2}\,\mathrm{m} = 1.44\,\mathrm{cm}}. \tag{33.1.4}$$

Because each color lands its fringes at its own spacing, white light spreads every order after the central one into a little spectrum, which is why the higher fringes show colored edges.

Checkpoint 33.1

In a double-slit experiment the fringe spacing is $\Delta y = \lambda L/d$. What happens to the spacing if (a) the slits are moved closer together, halving d; (b) red light is replaced by blue; (c) the whole apparatus is submerged in water, $n = 1.33$?

Answer

Answers to Checkpoints

Checkpoint 33.1. (a) Halving d doubles the spacing: slits closer together spread the pattern out. (b) Blue light has a shorter wavelength than red, so the spacing shrinks. (c) In water the wavelength shortens to λ/n, so every spacing shrinks by the factor 1.33: the fringes crowd together.

This page titled 33.1: Young's Double-Slit Experiment is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.2: Interference and Double-Slit Intensity

Equations (33.2) and (33.3) tell us *where* the bright and dark fringes are, but not how the brightness varies in between. For that we must add the two waves as waves, amplitude and phase together. The cleanest bookkeeping tool is the **phasor**: a rotating vector whose length is a wave's amplitude and whose angle is its phase. Adding two waves becomes adding two vectors, tip to tail, and "in phase" or "out of phase" becomes simple geometry.

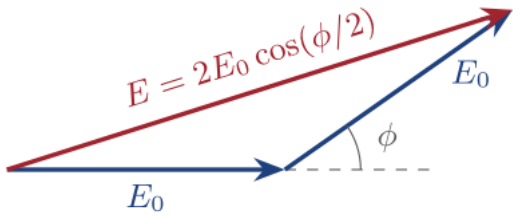


Figure 33.3. Phasor addition of two waves of equal amplitude E_0 differing in phase by ϕ. The resultant amplitude is $E = 2E_0\cos(\phi/2)$, largest when $\phi = 0$ (in phase) and zero when $\phi = \pi$ (opposed).

•Figure description

A phasor triangle. A blue arrow of length E nought points to the right from the origin. A second blue arrow of the same length E nought starts at its tip, rotated upward by the phase difference phi, marked by a small arc. A red arrow closes the figure from the origin to the far tip; it is the resultant, labeled E equals 2 E nought cosine of phi over 2. The triangle is isosceles, which is why the resultant takes that form.

The two waves arriving at P have equal amplitude E_0 but differ in phase by an amount set by their path difference. Since a full wavelength of path corresponds to 2π of phase, the **phase difference** is

$$\phi = \frac{2\pi}{\lambda}\,\delta = \frac{2\pi d}{\lambda}\sin\theta. \tag{33.6}$$

Drawn as phasors, the two equal vectors meet at angle ϕ (Figure 33.3). Their resultant, by the geometry of the isosceles triangle they form, has amplitude

$$E = 2E_0\cos\left(\frac{\phi}{2}\right). \tag{33.7}$$

Intensity is proportional to amplitude squared, so writing I_0 for the intensity a single slit would give and noting that two in-phase slits give $4I_0$ at the peak,

Key Equation: Double-slit intensity

$$I = 4I_0\cos^2\left(\frac{\phi}{2}\right), \qquad \phi = \frac{2\pi d}{\lambda}\sin\theta. \tag{33.8}$$

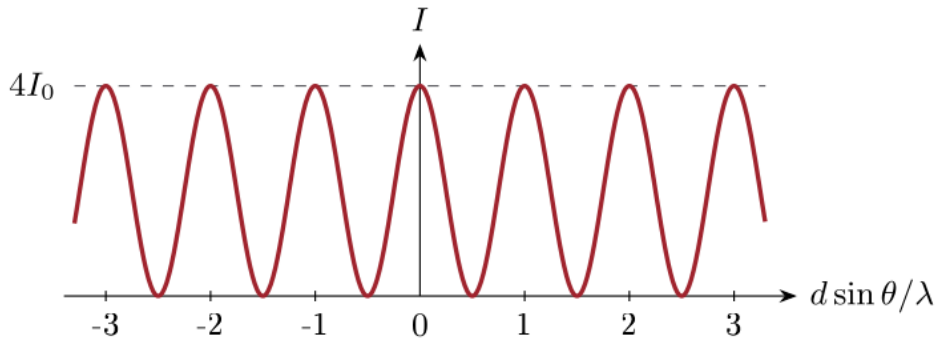


Figure 33.4. Double-slit intensity versus $d\sin\theta/\lambda$. Maxima of height $4I_0$ occur at integer orders m; minima (zero) fall halfway between. The brightness varies as $\cos^2$.

•Figure description

A graph of double-slit intensity against d sine theta over lambda. The curve is a smooth cosine-squared, rising to identical peaks of height 4 I nought at every integer value -- minus 3, minus 2, minus 1, 0, 1, 2, 3 -- and falling to exactly zero halfway between each pair. A dashed horizontal line at 4 I nought shows the peaks are all the same height, with no fading toward the edges.

This single expression contains the fringe conditions. The intensity peaks at $4I_0$ whenever ϕ is a multiple of 2π (that is, when $d\sin\theta = m\lambda$, recovering equation (33.2)) and drops to zero when ϕ is an odd multiple of π, i.e. $d\sin\theta = (m+\frac{1}{2})\lambda$, recovering equation (33.3). Between the peaks the brightness varies smoothly as $\cos^2$ (Figure 33.4). Notice the peaks are all the *same* height: idealized double-slit fringes do not fade with distance from center. (In a real experiment the finite width of each slit imposes a slowly falling envelope: the subject of the next chapter, diffraction.)

Example 33.3: Brightness between fringes

Two slits $0.180\ \text{mm}$ apart are lit by $656.3\ \text{nm}$ light, forming a pattern on a screen $80.0\ \text{cm}$ away. What fraction of the maximum intensity appears at a point $y = 0.600\ \text{cm}$ from the central maximum?

Solution

First find the phase difference there. With $\sin\theta \approx y/L = (0.600\times10^{-2})/(0.800) = 7.50\times10^{-3}$, the path difference is

$$\delta = d\sin\theta = (1.80\times10^{-4}\,\text{m})(7.50\times10^{-3}) = 1.35\times10^{-6}\,\text{m}, \tag{33.2.1}$$

so

$$\phi = \frac{2\pi}{\lambda}\delta = \frac{2\pi(1.35\times10^{-6})}{656.3\times10^{-9}} = 12.92\,\text{rad}. \tag{33.2.2}$$

Then, by equation (33.8),

$$\frac{I}{I_{\max}} = \cos^2\left(\frac{\phi}{2}\right) = \cos^2(6.462\,\text{rad}) = \boxed{0.968}. \tag{33.2.3}$$

The point lies close to a bright fringe ($\delta/\lambda = 2.06$, just past the second order), so it shines at about 97% of full brightness: consistent with sitting near, but not exactly on, the $m = 2$ maximum.

This page titled 33.2: Interference and Double-Slit Intensity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.3: Interference from Thin Films

The shimmer of color on a soap bubble or an oil slick is interference at work, with no slits at all. Light reflecting off the *top* of a thin transparent film interferes with light that enters, reflects off the *bottom*, and comes back out. Whether a given color reinforces or cancels depends on the film's thickness, so different thicknesses flash different colors.

Three things can shift the relative phase of two waves: traveling *different path lengths*, traveling through *different media* (where the wavelength shrinks to λ/n), and *reflection itself*. The third is easy to forget and crucial here: when light reflects from a medium of *higher* index of refraction, it flips phase by 180° (a jump of half a wavelength) just as a wave pulse on a rope inverts when it hits a fixed end. Reflecting from a *lower*-index medium causes no such flip.

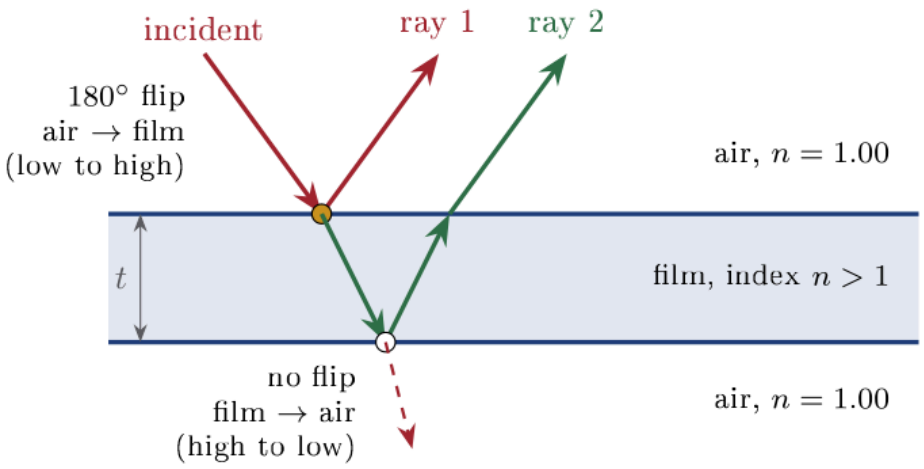


Figure 33.5. Interference in a thin film. Ray 1 reflects off the top surface, going from low to high index, and flips phase by 180° (filled dot). Ray 2 travels an extra path $2t$ inside the film and reflects off the bottom surface, going from high to low index, with no flip (open dot). The single flip is equivalent to an extra half-wavelength.

•Figure description

A thin film of index n and thickness t, drawn as three layers: air on top, the film in the middle, and air again below. A red incident ray comes down from the upper left onto the top surface. Ray 1 reflects straight back up; a gold dot where it reflects is labeled 180 degree flip, air to film, low to high index. The rest of the light, drawn in green as ray 2, refracts into the film and travels down to the bottom surface, where an open circle is labeled no flip, film to air, high to low index. Ray 2 reflects there, comes back up and out roughly parallel to ray 1, having traveled an extra distance of about 2 t inside the film. A thinner dashed ray continues on down into the lower air. The two emerging rays are what interfere.

Consider a film of index n and thickness t in air (Figure 33.5). Ray 1 reflects off the top surface (air-to-film, low-to-high) and *does* flip. Ray 2 passes into the film, reflects off the bottom surface (film-to-air, high-to-low) and does *not* flip, then exits having traveled an extra geometric distance $2t$ inside the film. Two effects therefore separate the rays: the half-wavelength from the single reflection flip, and the path $2t$, which holds $2nt/\lambda$ wavelengths because the wavelength inside the film is λ/n. With exactly one flip, the conditions for the *reflected* light are

Key Equation: Thin film, one reflection phase flip

$$2nt = \left(m + \tfrac{1}{2}\right)\lambda, \qquad \text{(constructive, bright reflection)} \tag{33.9}$$

$$2nt = m\lambda, \qquad \text{(destructive, dark reflection)} \tag{33.10}$$

with $m = 0, 1, 2, \ldots$ and λ the wavelength in air.

Example 33.4: A soap film at its thinnest

A soap film ($n = 1.33$) in air is lit by light of free-space wavelength $600\ \text{nm}$. Find the minimum nonzero thickness that makes the reflected light interfere constructively (so the film looks bright at that color).

Solution

A soap film is air on both sides, so there is exactly one reflection flip (at the top). Constructive reflection therefore obeys equation (33.9), $2nt = (m + \frac{1}{2})\lambda$. The thinnest film takes $m = 0$:

$$t = \frac{\lambda}{4n} = \frac{600\,\text{nm}}{4(1.33)} = \boxed{113\,\text{nm}}. \tag{33.3.1}$$

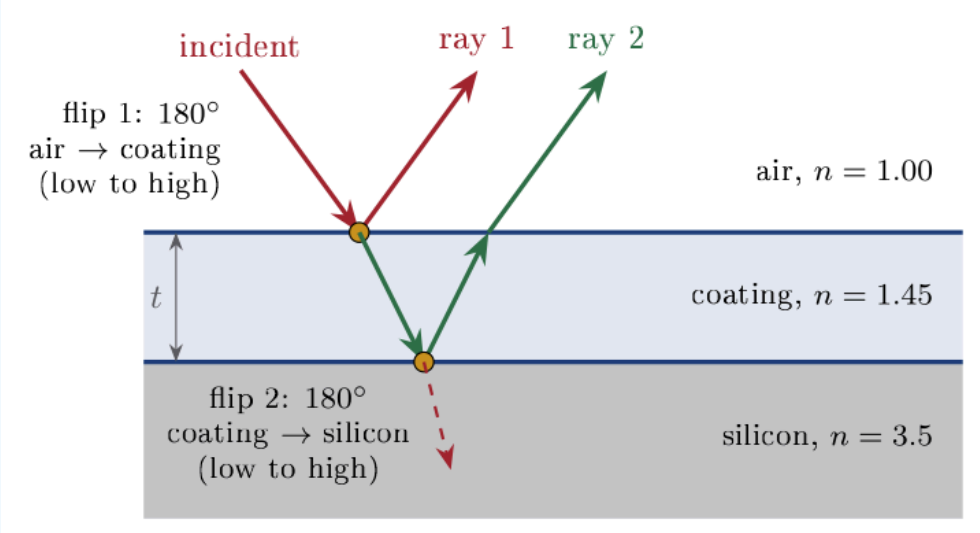
A bubble's wall thins continuously as it drains under gravity; as t sweeps through the values that satisfy this condition for each color in turn, the wall flashes through the spectrum, and just before it pops, where $t \to 0$ and $2nt \to 0$ (the $m = 0$ *destructive* case), it goes black.

Example 33.5: An anti-reflection coating

A silicon solar cell ($n = 3.5$) is coated with a thin film of silicon dioxide ($n = 1.45$) to cut reflective losses. Find the minimum film thickness that gives the *least* reflection at $\lambda = 550\ \text{nm}$.

Solution

Here the geometry is different: light goes air ($n = 1$) $\to$ film ($n = 1.45$) $\to$ silicon ($n = 3.5$). *Both* reflections are now low-to-high, so *both* rays flip by $180°$, as the sketch shows. Compare Figure 33.5, where the bottom reflection is high-to-low and does not flip.



•Figure description
An anti-reflection coating, drawn as three layers: air on top, a thin coating of index 1.45 and thickness t in the middle, and silicon of index 3.5 at the bottom. An incident ray comes down from the upper left. Ray 1 reflects off the top of the coating; a gold dot there is labeled flip 1, 180 degrees, air to coating, low to high. Ray 2, in green, refracts into the coating, reflects off the silicon, where a second gold dot is labeled flip 2, 180 degrees, coating to silicon, low to high, and comes back out parallel to ray 1. A thinner dashed ray continues on into the silicon. Because both reflections flip, the flips cancel and only the extra path 2 n t inside the coating separates the two rays.

Two flips cancel (there is *no net* reflection phase shift) so the conditions are the reverse of the single-flip case: least reflection (destructive) now requires a half-integer path,

$$2nt = \left(m + \tfrac{1}{2}\right)\lambda. \tag{33.3.2}$$

The minimum thickness is again $m = 0$:

$$t = \frac{\lambda}{4n} = \frac{550\,\text{nm}}{4(1.45)} = \boxed{94.8\,\text{nm}}. \tag{33.3.3}$$

A coating one quarter-wavelength thick (measured inside the film) makes the two reflected rays cancel, so almost no light bounces back and more passes into the cell. The same quarter-wave trick coats camera lenses and eyeglasses: the faint purple sheen you see on good optics.

The two examples carry a warning worth stating plainly: the boxed conditions, Equations (33.9)–(33.10), assume *one* reflection flip. Always count the flips first: at each surface, ask whether the light is going from lower to higher index (a flip) or from higher to lower (no flip). An *odd* number of low-to-high reflections gives a net half-wave shift and the conditions above; an *even* number (zero or two) gives no net shift, and constructive and destructive swap places.

Checkpoint 33.2

(a) A vertical soap film in air drains and thins toward the top. Just before it pops, the thinnest region looks black by reflected light. Why black rather than bright? (b) An antireflection coating on a lens has a refractive index between those of air and glass. How many reflection phase flips occur, and which condition (equation (33.9) or (33.10)) now gives the dark reflection?

Answer

Answers to Checkpoints

Checkpoint 33.2. (a) As the film's thickness t approaches zero, the path difference $2nt$ approaches zero too. For a soap film in air there is exactly one phase flip, at the front (air-to-soap) surface, so with zero path difference the two reflected waves are exactly out of step and cancel. The thinnest part reflects no light and looks black. (b) Two. The light flips on reflecting from the air–coating surface *and* again from the coating–glass surface, since each goes from lower to higher index. The two flips cancel, so the conditions swap: the dark reflection now occurs when $2nt = (m + \frac{1}{2})\lambda$, the form of equation (33.9). The thinnest such coating is a quarter-wavelength thick.

This page titled 33.3: Interference from Thin Films is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.4: Interferometry

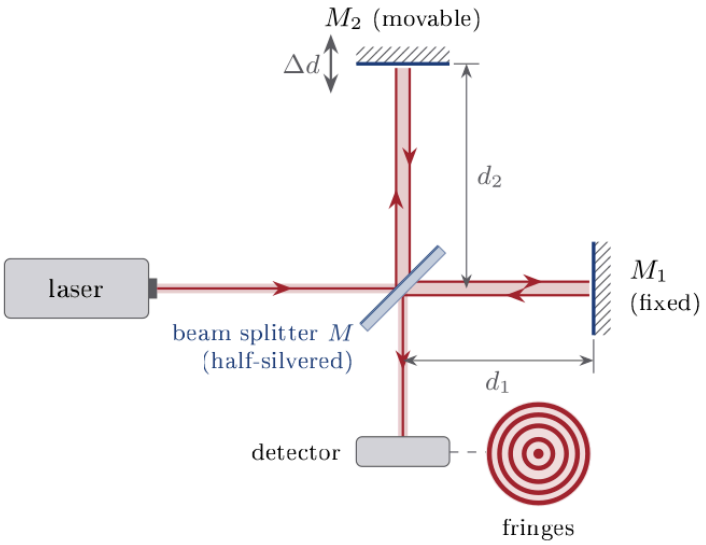


Figure 33.6. The Michelson interferometer. The half-silvered beam splitter M divides the laser light between a fixed arm (d_1) and a movable arm (d_2); each beam reflects back, and the recombined beams interfere at the detector, often as a set of rings. The path difference is $2d_2 - 2d_1$, so sliding M_2 by Δd changes it by $2\ \Delta d$.

•Figure description

A Michelson interferometer. A laser at the left sends a red beam to the right into a beam splitter, a thin half-silvered glass plate tilted at 45 degrees. Half the light passes straight through along an arm of length d sub 1 to a fixed mirror M sub 1 and reflects back; the other half turns upward along an arm of length d sub 2 to a movable mirror M sub 2 and reflects back. Small arrowheads on each arm show the light going out and coming back. The mirrors are drawn with hatched backs, and a double arrow labeled delta d shows that M sub 2 slides up and down. The two returning beams recombine at the splitter and travel down together to a detector, beside which an inset shows the resulting interference pattern as concentric red and white rings. The path difference is twice d sub 2 minus twice d sub 1.

Because interference responds to path differences as small as a fraction of a wavelength, it is the basis of the most precise length measurements we can make. The classic instrument is the **Michelson interferometer** (Figure 33.6). A beam from a single source strikes a half-silvered mirror M (the **beam splitter**), which sends half the light down one arm to a fixed mirror M_1 and half down a perpendicular arm to a movable mirror M_2. Both beams reflect back, recombine at M, and travel together to the detector, where they interfere.

The round trip down each arm doubles its length, so the path difference between the two beams is $2d_2 - 2d_1$. Now slide the movable mirror by a distance Δd. The path in that arm changes by $2\ \Delta d$, and each time this change adds one more wavelength, the pattern at the detector advances by exactly one fringe: one full cycle from bright to bright. Counting N fringes as the mirror moves gives

Key Equation: Michelson interferometer

$$2\,\Delta d = N\lambda. \tag{33.11}$$

Because N is a pure count of fringes (an integer you can tally exactly) and λ is tiny, this turns a delicate length into a precise number. Move the mirror a known distance and you measure λ; know λ and you measure the displacement to a fraction of a wavelength. The same principle, scaled to arms four kilometers long and mirrors hung as pendulums, is how the LIGO observatories detected gravitational waves: a passing ripple in spacetime stretched one arm relative to the other by less than the width of a proton, and the fringe count registered it.

Example 33.6: Counting fringes

In a Michelson interferometer using $632.8\ \text{nm}$ laser light, the movable mirror is advanced slowly and 500 bright fringes pass the detector. How far did the mirror move?

Solution

From equation (33.11) with $N = 500$,

$$\Delta d = \frac{N\lambda}{2} = \frac{500\,(632.8 \times 10^{-9}\,\text{m})}{2} = \boxed{1.58 \times 10^{-4}\,\text{m} = 0.158\,\text{mm}}. \tag{33.4.1}$$

A sixth of a millimeter, measured by counting: no ruler involved. This is why interferometers are used to calibrate length standards.

Checkpoint 33.3

A Michelson interferometer uses light of wavelength $500\ \text{nm}$. (a) One mirror is moved $0.25\ \mu\text{m}$. How many bright fringes pass the center of the field? (b) Why does equation (33.11) contain $2\Delta d$ rather than Δd?

Answer

Answers to Checkpoints

Checkpoint 33.3. (a) $N = 2\Delta d/\lambda = 2(0.25\ \mu\text{m})/(0.500\ \mu\text{m}) = 1$: exactly one fringe. (b) The light travels to the moving mirror and back, so moving the mirror by Δd lengthens that arm's round trip by $2\Delta d$. A new bright fringe appears each time the round-trip path changes by one wavelength, that is, each time the mirror moves half a wavelength.

This page titled 33.4: Interferometry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.S: Interference (Summary)

Summary

- Light from narrow slits spreads by Huygens' principle (Chapter 31) and overlaps. Visible interference needs **coherent** (fixed-phase) light, which is why a single source illuminates both slits.
- **Young's double slit**: path difference $\delta = d\sin\theta$. Bright fringes at $d\sin\theta = m\lambda$, dark at $d\sin\theta = (m+\frac{1}{2})\lambda$. Near the center, bright fringes lie at $y = m\lambda L/d$ with spacing $\Delta y = \lambda L/d$.
- **Intensity**: adding two waves as phasors gives $I = 4I_0\cos^2(\phi/2)$ with phase difference $\phi = (2\pi d/\lambda)\sin\theta$. Maxima ($4I_0$) at integer orders, zeros halfway between.
- **Thin films**: reflection from a higher-index medium flips phase by 180° ($\lambda/2$). For one net flip, reflected light is bright when $2nt = (m+\frac{1}{2})\lambda$ and dark when $2nt = m\lambda$. Count the flips: an even number swaps the conditions. A quarter-wave coating ($t = \lambda/4n$) minimizes reflection.
- **Michelson interferometer**: moving a mirror by Δd shifts N fringes, with $2\ \Delta d = N\lambda$, the basis of ultra-precise length and wavelength measurement.

Key Equations

Double-slit path difference

$$\delta = d\sin\theta \tag{33.S.1}$$

Double-slit fringe conditions

$$\begin{aligned} d\sin\theta &= m\lambda && \text{(bright)} \\ d\sin\theta &= \left(m+\tfrac{1}{2}\right)\lambda && \text{(dark)} \end{aligned} \tag{33.S.2}$$

Fringe positions and spacing (small angles)

$$y_{\text{bright}} = m\frac{\lambda L}{d}, \qquad \Delta y = \frac{\lambda L}{d} \tag{33.S.3}$$

Phase difference from path difference

$$\phi = \frac{2\pi}{\lambda}\delta = \frac{2\pi d}{\lambda}\sin\theta \tag{33.S.4}$$

Phasor sum of two equal waves

$$E = 2E_0\cos\left(\frac{\phi}{2}\right) \tag{33.S.5}$$

Double-slit intensity

$$I = 4I_0\cos^2\left(\frac{\phi}{2}\right) \tag{33.S.6}$$

Wavelength inside a medium

$$\lambda_n = \frac{\lambda}{n} \tag{33.S.7}$$

Thin film, one reflection phase flip

$$\begin{aligned} 2nt &= \left(m + \tfrac{1}{2}\right)\lambda \quad && \text{(bright reflection)} \\ 2nt &= m\lambda && \text{(dark reflection)} \end{aligned} \tag{33.S.8}$$

Quarter-wave coating

$$t = \frac{\lambda}{4n} \tag{33.S.9}$$

Michelson interferometer

$$2\,\Delta d = N\lambda \tag{33.S.10}$$

This page titled 33.S: Interference (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

33.E: Interference (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Take the speed of light as $c = 3.00 \times 10^8\ \text{m/s}$ and, where needed, treat angles as small ($\sin\theta \approx \tan\theta = y/L$).

Section 33.1 — Young's Double-Slit Experiment

33.1 ★★ *[Conceptual.]* Two flashlights shine on the same spot on a wall, yet no interference fringes appear. Two slits lit by one laser do show fringes. Explain the difference in terms of coherence.

33.2 ★ Light of wavelength $550\ \text{nm}$ falls on two slits $0.10\ \text{mm}$ apart. At what angle does the first-order ($m = 1$) bright fringe appear?

33.3 ★★ A double slit with $d = 0.040\ \text{mm}$ is lit by $633\ \text{nm}$ laser light, with the screen $2.0\ \text{m}$ away. Find the spacing between adjacent bright fringes.

33.4 ★★ In a double-slit experiment the third-order bright fringe is found $2.7\ \text{cm}$ from the center on a screen $1.8\ \text{m}$ away, with slits $0.15\ \text{mm}$ apart. Find the wavelength.

33.5 ★★★ *[Real-world application: metrology.]* A double slit is illuminated simultaneously by $480\ \text{nm}$ and $600\ \text{nm}$ light ($d = 0.030\ \text{mm}$, $L = 1.2\ \text{m}$). (a) Find the fringe spacing for each color. (b) At what distance from center do a bright fringe of one color and a bright fringe of the other first coincide?

33.6 ★★ In a double-slit experiment, slits $0.20\ \text{mm}$ apart are lit by $550\ \text{nm}$ light, with the screen $1.8\ \text{m}$ away. (a) Find the spacing between adjacent bright fringes. (b) How far from the center is the third bright fringe?

33.7 ★★ Two slits $0.15\ \text{mm}$ apart produce a pattern on a screen $2.0\ \text{m}$ away. The fourth bright fringe lies $2.9\ \text{cm}$ from the central maximum. Find the wavelength of the light.

Section 33.2 — Interference and Double-Slit Intensity

33.8 ★★ At a certain point in a double-slit pattern the phase difference between the two waves is $\phi = \pi/3$. What fraction of the maximum intensity appears there?

33.9 ★★ Two slits $0.20\ \text{mm}$ apart are lit by $589\ \text{nm}$ light on a screen $1.0\ \text{m}$ away. Find the intensity (as a fraction of maximum) at a point $1.5\ \text{mm}$ from the central maximum.

33.10 ★★★ *[Conceptual.]* The idealized double-slit pattern has every maximum the same height, $4I_0$. Explain why a single slit gives intensity I_0 but two coherent slits give $4I_0$, not $2I_0$, at the center, and reconcile this with conservation of energy.

33.11 ★★ In a two-slit interference pattern the central intensity is $I_{\max}$. At a point where the path difference between the two waves is $\lambda/4$, what fraction of $I_{\max}$ is the intensity?

Section 33.3 — Interference from Thin Films

33.12 ★ Light reflects from the top of a water film ($n = 1.33$) sitting on glass. At the air-water surface, does the reflected wave undergo a 180° phase flip? Explain.

33.13 ★★ A soap film ($n = 1.33$) in air appears brightest in reflection at $\lambda = 540\ \text{nm}$. Find the minimum nonzero film thickness.

33.14 ★★★ *[Real-world application: optics.]* A camera lens ($n = 1.52$) is coated with magnesium fluoride ($n = 1.38$) to reduce reflection at $\lambda = 550\ \text{nm}$. (a) How many reflection phase flips occur? (b) Find the minimum coating thickness for least reflection, and state whether the condition is the same as or opposite to a soap film's.

33.15 ★★ A soap film ($n = 1.33$) surrounded by air looks bright green ($\lambda = 510\ \text{nm}$) in reflected light. Find the two smallest film thicknesses that give this constructive reflection. (One phase reversal occurs, at the front surface.)

Section 33.4 — Interferometry & the Double-Slit Puzzle

33.16 ★★ A Michelson interferometer uses $589\ \text{nm}$ light. The movable mirror is shifted until 250 fringes pass. How far did it move?

33.17 ★★★ *[Real-world application: precision measurement.]* A Michelson interferometer's movable mirror is advanced 0.233 mm, and 750 fringes are counted. Find the wavelength of the light used.

33.18 ★★★ *[Conceptual — the double-slit puzzle.]* Return to the chapter opener. Young's experiment shows light behaving as a wave; the same fringes appear when the experiment is run with single electrons sent through one at a time.

a. If electrons are sent one at a time, each makes a single dot on the screen, yet over many electrons an interference pattern builds up. What does this imply about a single electron's passage through the slits?
b. If a detector is placed at the slits to record which slit each electron goes through, the fringes vanish and two simple bands appear. In one or two sentences, state what this says about the role of measurement, connecting it to the opening question about quantum mechanics.

Challenge Problems

33.19 ★★★ *[Real-world application: metrology — the optical wedge.]* Two flat glass plates $12.0\ \text{cm}$ long are laid one on the other and touch along one edge. A thin wire laid across the far edge props them apart, leaving a wedge of air between them. Viewed from above in reflected $589\ \text{nm}$ light, the wedge shows a series of straight, evenly spaced dark fringes. Recall that of the two reflections, from the bottom of the upper plate and from the top of the lower plate, only one suffers a 180° phase flip.

a. Explain why the line of contact is *dark*, not bright, and show that dark fringes occur where the air gap has thickness $t = m\lambda/2$.
b. If the gap grows linearly from 0 at the contact line to D at the wire, show that the number of dark fringes across the full length is $N = 2D/\lambda$, and that the fringe spacing is $\Delta x = \lambda\ell/(2D)$.
c. An observer counts 85 dark fringes across the $12.0\ \text{cm}$. Find the wire's diameter and the fringe spacing.
d. The wedge is now filled with water ($n = 1.33$) instead of air. How many fringes appear, and does the technique become more or less sensitive? Explain why this method can measure a thickness far smaller than the wavelength of the light used.

33.20 ★★★ *[Real-world application: radio astronomy — the sea-cliff interferometer.]* A radio telescope sits on a cliff of height h above a calm sea, watching a radio source rise above the horizon at angle θ. The antenna receives two waves: one direct from the source, and one reflected from the sea surface. The reflection off the conducting seawater introduces a 180° phase flip. (This arrangement, used in Australia in the 1940s, gave the resolution of a two-dish interferometer with only one dish.)

a. Treating the reflected wave as coming from an image source a distance h *below* the sea surface, show that the path difference between the two waves is $\delta = 2h\sin\theta$.
b. Including the phase flip, show that maxima occur where $2h\sin\theta = \left(m + \frac{1}{2}\right)\lambda$ and minima where $2h\sin\theta = m\lambda$.
c. For $h = 25\ \text{m}$ and $\lambda = 1.5\ \text{m}$, find the angle of the first maximum above the horizon, and the angular spacing between successive maxima at small θ.
d. The effective baseline of this instrument is $2h$, not h. Explain why, and find the angular resolution $\lambda/2h$ in degrees. What practical limitation would you expect on a day when the sea is rough?

33.21 ★★★ *[Real-world application: precision measurement — measuring the refractive index of air.]* One arm of a Michelson interferometer contains an evacuated glass cell of length $L = 5.00\ \text{cm}$ with thin, flat windows. The interferometer is illuminated with $633\ \text{nm}$ helium–neon laser light. Air is slowly admitted to the cell, and the fringes are counted as they sweep past.

a. The light crosses the cell *twice*. Show that filling the cell with a gas of index n shifts the fringe count by $N = 2L(n-1)/\lambda$.
b. 46 fringes are counted between full vacuum and atmospheric pressure. Find n for air, to six decimal places.
c. For a dilute gas, $n - 1$ is proportional to pressure. If the observer can reliably detect a shift of a quarter of a fringe, what is the smallest pressure change (as a fraction of atmospheric, and in pascals) that this apparatus can sense?
d. Suggest one change to the apparatus that would improve the sensitivity of part (c) by a factor of ten, and state the practical cost of making it.

This page titled 33.E: Interference (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34: DIFFRACTION

In the last chapter, two slits gave us a row of interference fringes. It is natural to push on the idea: if two slits make a pattern, what does *one* slit do? What about five slits, or a million? The answer is that a single slit makes a pattern of its own (a bright central band flanked by weaker fringes) and that adding more slits does not blur this into a wash of light but sharpens it into a row of needle-thin lines. The same physics, scaled from one opening to millions, runs from the blur at the edge of a shadow all the way to the instrument that first revealed the spacing of atoms in a crystal.

This spreading of waves as they pass an aperture or an edge is **diffraction**, and it is not a separate phenomenon from interference: it *is* interference, among the infinitely many Huygens wavelets that fan out from across a single opening. Diffraction sets a hard limit on the sharpness of every image ever formed: it is why a telescope, however perfect its mirror, cannot resolve two stars too close together, and why your own eye has a finite acuity. It also gives us the most powerful tools we have for taking light apart by wavelength and for reading the atomic architecture of matter. We will work from the single slit outward, to its intensity pattern, to circular apertures and the resolution of instruments, to the double slit seen afresh, to the diffraction grating, and finally to X-rays scattering from the planes of a crystal.

Learning Objectives

- Locate the minima of a single-slit diffraction pattern and find the width of the central maximum.
- Compute the intensity at any angle in single-slit diffraction.
- Apply the resolution limit of a circular aperture and Rayleigh's criterion.
- Describe how single-slit diffraction modulates the double-slit interference pattern.
- Use the grating equation and understand why many slits give sharp, bright lines.
- Apply Bragg's law to the diffraction of X-rays by crystals.

We will return to the opening question (one slit, many slits) as the through-line of the chapter.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

34.1: Single-Slit Diffraction
34.2: Diffraction by a Circular Aperture
34.3: Diffraction by a Double Slit
34.4: Diffraction Gratings
34.5: Diffraction of X-Rays by Crystals
34.S: Diffraction (Summary)
34.E: Diffraction (Problems)

This page titled 34: Diffraction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.1: Single-Slit Diffraction

Recall Huygens' principle from Chapter 31: every point on a wavefront is itself a source of spherical wavelets. A slit is not infinitely narrow; it has a width a, and every point *across* that width radiates. The light reaching a point on a distant screen is the sum of all these wavelets, and because they set out from different parts of the slit they arrive having traveled slightly different distances. Their interference (now among a continuum of sources rather than just two) produces the single-slit pattern.

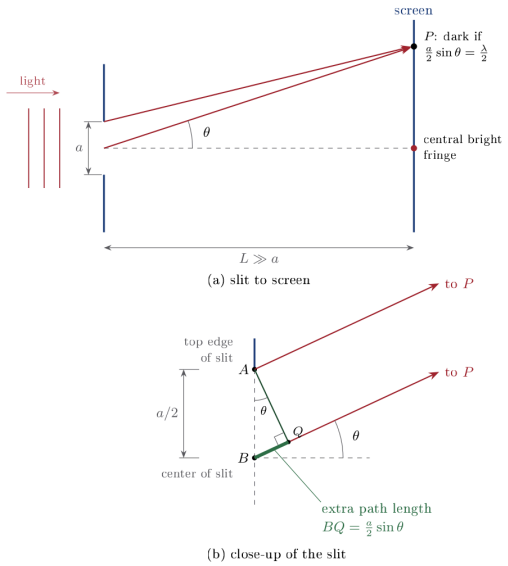


Figure 34.1. Single-slit diffraction by the pairing argument. (a) Rays from the top edge A and the center B of the slit travel to the same point P on a distant screen. (b) Close-up of the slit. Because $L \gg a$, the two rays are nearly parallel, and the perpendicular AQ marks where they are in step, so the lower ray travels an extra distance BQ. In the right triangle ABQ the hypotenuse is $a/2$ and the angle at A equals θ, so $BQ = (a/2)\sin\theta$. When this is $\lambda/2$, every such pair cancels and P is dark: the first minimum, $a\sin\theta = \lambda$.

•Figure description

Single-slit diffraction by the pairing argument, in two panels. Panel a shows the whole set-up: plane waves arrive from the left at a slit of width a, and a screen stands a distance L to the right, with L much larger than a. A dashed axis runs from the center of the slit to the central bright fringe on the screen. Two red rays, one from the top edge of the slit and one from its center, run up to the same point P on the screen, at angle theta above the axis; P is labeled dark if a over 2 times sine theta equals lambda over 2. Panel b is the close-up. Point A is the top edge of the slit and point B is the center, a distance a over 2 below it. The two rays toward P leave A and B parallel, at angle theta above the horizontal. A line from A meets the lower ray at a right angle at point Q. This makes a right triangle A B Q with hypotenuse A B equal to a over 2 and angle theta at A, so the short side B Q, highlighted in green as the extra path of the lower ray, equals a over 2 times sine theta.

The dark fringes are the easy part, and a clever pairing argument finds them exactly. Imagine dividing the slit into an upper half and a lower half (Figure 34.1). Pair the ray from the top edge of the slit (point A) with the ray from the top of the lower half (point B, the center of the slit): two rays that start $a/2$ apart and head for the same point P on the screen. Because the screen is far away, the two rays are very nearly parallel. The close-up in Figure 34.1(b) shows the geometry. Drop a perpendicular from A to the lower ray, meeting it at Q; from the line AQ onward the two rays travel equal distances to P, so the only difference is the short stretch BQ that the lower ray covers first. That stretch is a side of the right triangle ABQ, whose hypotenuse is $AB = a/2$ and whose angle at A equals θ (the same angle the rays make with the horizontal), so the path difference is

$$BQ = \frac{a}{2}\sin\theta. \tag{34.1.1}$$

If this equals exactly half a wavelength, the two wavelets arrive crest to trough and cancel. The same is true for every other pair spaced $a/2$ apart: a point just below A with the point just below B, and so on down the slit, until every point of the upper half has a partner in the lower half. Every wavelet from the slit is canceled by its partner, so no light at all arrives at P: the screen is dark there. Setting $(a/2)\sin\theta = \lambda/2$ gives the first minimum at $a\sin\theta = \lambda$.

The other dark fringes come from the same trick with smaller pieces. Divide the slit into four equal strips, each $a/4$ wide. Pair

each point of strip 1 with the point $a/4$ below it in strip 2, and do the same for strips 3 and 4. These pairs cancel when $(a/4)\sin\theta = \lambda/2$, that is, when $a\sin\theta = 2\lambda$: the second dark fringe. Six strips of width $a/6$ cancel in pairs when $(a/6)\sin\theta = \lambda/2$, giving $a\sin\theta = 3\lambda$, and so on. Splitting the slit into $2m$ strips gives the mth dark fringe:

Key Equation: Single-slit minima

$$a\sin\theta = m\lambda, \qquad m = 1, 2, 3, \ldots \tag{34.1}$$

Two features deserve emphasis. First, equation (34.1) looks like the double-slit *bright*-fringe condition from Chapter 33, but here it gives the *dark* fringes, and there is no $m = 0$; the center is the brightest point, not a minimum. Second, the central maximum is twice as wide as the others: it stretches from the $m = -1$ minimum to the $m = +1$ minimum. On a screen a distance L away, with small angles ($\sin\theta \approx y/L$), the first minimum sits at $y_1 = \lambda L/a$, so the central bright band has full width

$$w = 2y_1 = \frac{2\lambda L}{a}. \tag{34.2}$$

The dark fringes are the landmarks of the whole pattern (Figure 34.2). For small angles they fall at evenly spaced positions $y_m = m\lambda L/a$, and each bright fringe simply fills the gap between two neighboring dark fringes. The central maximum spans the gap from $-y_1$ to $+y_1$, a width of $2\lambda L/a$; every side maximum spans a single gap, from y_m to y_{m+1}, a width of only $\lambda L/a$, with its peak near the middle of the gap. So finding the dark fringes is all it takes to locate the bright ones and measure their widths.

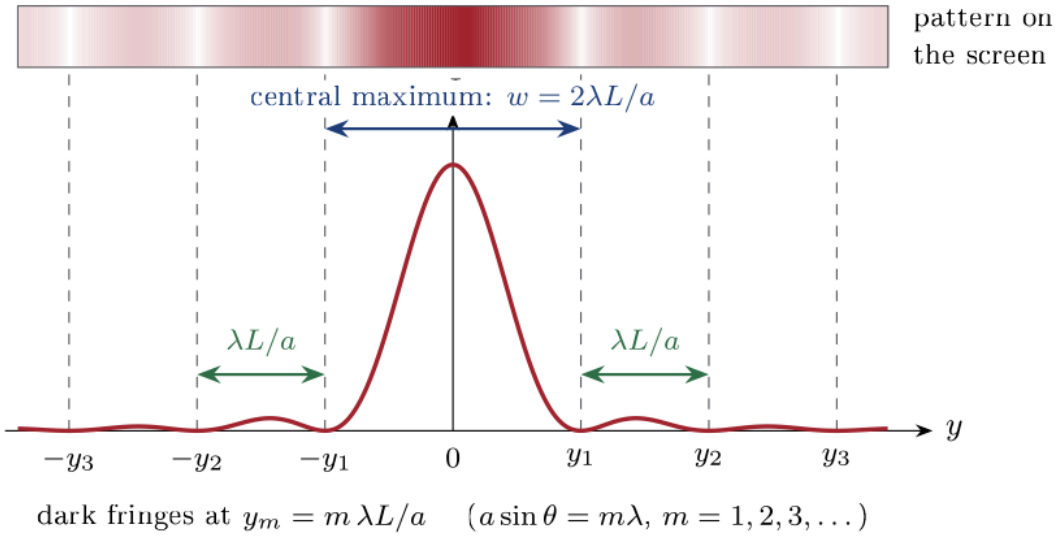


Figure 34.2. Reading the bright fringes from the dark ones. For small angles the dark fringes fall at evenly spaced positions $y_m = m\lambda L/a$ on the screen. Each bright fringe fills the gap between two neighboring dark fringes: the central maximum spans $-y_1$ to $+y_1$ (width $2\lambda L/a$), and each side maximum spans one gap (width $\lambda L/a$), with its peak near the middle. The shading of the strip is exaggerated so that the faint side fringes show.

•Figure description

A graph of single-slit intensity I against position y on the screen, with a strip above it showing the pattern as it looks on the screen: a wide bright red band in the middle and faint narrower bands on each side, the shading exaggerated so the faint bands show. Dashed vertical lines mark the dark fringes at y sub 1, y sub 2 and y sub 3 on each side of center, evenly spaced, and run up through the strip so each dark line on the graph lines up with a dark gap in the strip. The intensity curve is a tall peak centerd at zero that falls to zero at minus y sub 1 and plus y sub 1, followed by small humps between successive dark fringes. A blue double arrow spans the central maximum from minus y sub 1 to plus y sub 1, labeled central maximum, w equals 2 lambda L over a. Green double arrows span the gaps from y sub 1 to y sub 2 and from minus y sub 2 to minus y sub 1, each labeled lambda L over a. A note below reads: dark fringes at y sub m equals m lambda L over a, where a sine theta equals m lambda, m equals 1, 2, 3 and so on.

Notice the slit width a is in the *denominator*: the *narrower* the slit, the *wider* the pattern spreads. Squeezing a wave makes it fan out: the hallmark of diffraction.

The same factor explains why diffraction escapes notice in everyday life. Light passing through a doorway $1\ \text{m}$ wide has $a/\lambda \approx 2$

$\times 10^{6}$, so its first minimum lies only $\theta \approx \lambda/a \approx 5 \times 10^{-7}\ \text{rad}$ from the straight-ahead direction. Across a $5\ \text{m}$ room that is a few micrometers: far too small to see, so the doorway seems to cast a perfectly sharp shadow. Sound is another story. A $340\ \text{Hz}$ note has a wavelength of about $1\ \text{m}$, comparable to the doorway, so sound spreads widely as it passes through. That is why you can hear someone around a corner but cannot see them.

Example 34.1: Width of the central maximum

Light of wavelength $580\ \text{nm}$ falls on a slit $0.300\ \text{mm}$ wide; the screen is $2.00\ \text{m}$ away. (a) Find the width of the central bright fringe. (b) What happens if the slit is widened tenfold, to $3.00\ \text{mm}$?

Solution

(a) The first minimum is at $y_1 = \lambda L/a$, and the central maximum spans from $-y_1$ to $+y_1$:

$$w = \frac{2\lambda L}{a} = \frac{2(580\times10^{-9}\,\text{m})(2.00\,\text{m})}{3.00\times10^{-4}\,\text{m}} = \boxed{7.73\times10^{-3}\,\text{m} = 7.73\,\text{mm}}. \tag{34.1.2}$$

(b) Since $w \propto 1/a$, widening the slit by a factor of ten *narrows* the central band by a factor of ten, to $0.773\ \text{mm}$. As the slit grows large compared with the wavelength, the pattern collapses toward the sharp bright rectangle of geometric optics: diffraction becomes negligible, which is why we rarely notice it for everyday openings.

Intensity in Single-Slit Diffraction

To find not just where the dark fringes are but how the brightness varies everywhere, we again add the wavelets as phasors, but now there is a continuous fan of them, one from each point across the slit, each a hair more retarded in phase than the last. Adding a continuum of equal phasors that turn through a steadily increasing angle traces an arc of a circle; the resultant is the chord across that arc. Carrying out the sum gives the intensity as a function of angle:

Key Equation: Single-slit intensity

$$I(\theta) = I_{\text{max}}\left(\frac{\sin\alpha}{\alpha}\right)^2, \qquad \alpha = \frac{\pi a}{\lambda}\sin\theta. \tag{34.3}$$

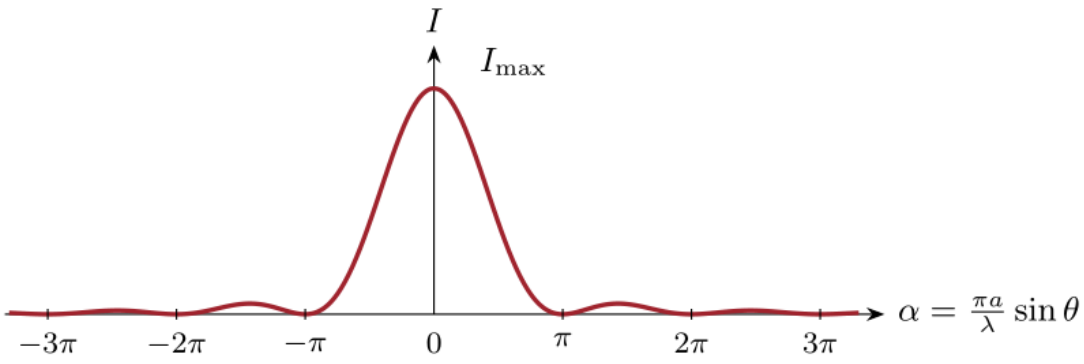


Figure 34.3. Single-slit intensity $(\sin\alpha/\alpha)^2$. The broad central maximum carries almost all the light; minima fall at $\alpha = m\pi$, i.e. $a\sin\theta = m\lambda$. Secondary maxima are faint.

•Figure description

A graph of single-slit intensity against alpha, which equals pi a over lambda times sine theta. The curve is the function sine alpha over alpha, squared. It has one tall broad central peak of height I max at alpha equals zero, dropping to exactly zero at alpha equals plus and minus pi, 2 pi and 3 pi. Between those zeros are secondary maxima that are very much smaller -- the first is under five per cent of the central height -- and they shrink further outward. Almost all the light sits in the central peak.

Here I_{max} is the intensity at the center, and α is half the phase difference between the wavelets from the two *edges* of the slit. The function $(\sin\alpha/\alpha)^2$ is worth reading closely (Figure 34.3). At the center $\theta = 0$, $\alpha \to 0$, and $\sin\alpha/\alpha \to 1$: full brightness. The intensity falls to zero whenever $\sin\alpha = 0$ with $\alpha \neq 0$, that is, $\alpha = m\pi$, which is precisely $a\sin\theta = m\lambda$, recovering the minima of equation

(34.1). Between the zeros sit weak **secondary maxima**, the first of which carries less than 5% of the central intensity. Almost all the light is in the central peak.

Checkpoint 34.1

(a) A single slit is narrowed to half its original width. What happens to the angular width of the central bright fringe? (b) What happens if the slit is made narrower than the wavelength of the light? *(Hint: look at the condition for the first minimum,* $a \sin\theta = \lambda$*.)*

Answer

Answers to Checkpoints

Checkpoint 34.1. (a) The first minimum is at $\sin\theta = \lambda/a$, so halving a doubles $\sin\theta$ (and, for small angles, doubles the angle): the central fringe becomes about twice as wide. A narrower slit spreads the light *more*: the opposite of what ray optics would suggest. (b) If $a < \lambda$, the condition $\sin\theta = \lambda/a$ would require $\sin\theta > 1$, which is impossible. There are no minima at all: the central maximum fills the entire half-space beyond the slit, and the slit acts like a point source.

This page titled 34.1: Single-Slit Diffraction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.2: Diffraction by a Circular Aperture

Lenses, mirrors, telescopes, and eyes gather light through *round* openings, not slits, so their diffraction pattern is a set of concentric rings: a bright central disk (the **Airy disk**) ringed by faint circles. The mathematics is harder, but the result differs from the slit only by a numerical factor: the first dark ring of a circular aperture of diameter D lies at

Key Equation: Circular-aperture first minimum

$$\sin\theta = 1.22\,\frac{\lambda}{D}. \tag{34.4}$$

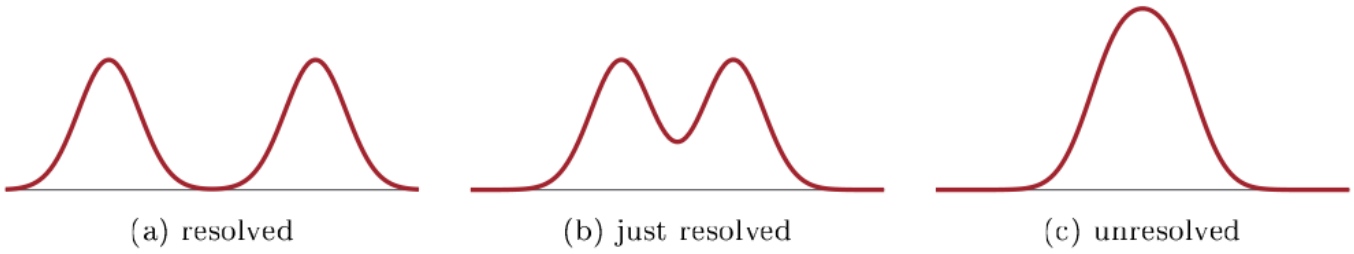


Figure 34.4. Rayleigh's criterion. Two point sources each make a diffraction pattern. (a) Well separated, they show two clear peaks. (b) At the Rayleigh limit, one peak sits on the other's first minimum: *just* resolved, a shallow dip remains. (c) Closer still, the patterns merge into a single blur and the sources cannot be told apart.

•Figure description

Three side-by-side intensity graphs showing two point sources at decreasing separation. In panel (a), resolved, the two peaks stand well apart with a deep valley between them. In panel (b), just resolved, the peaks have moved close enough that one sits on the other's first minimum; a shallow dip still separates them, which is the Rayleigh limit. In panel (c), unresolved, the peaks have merged into a single broad hump with no dip at all, and the two sources can no longer be told apart.

The factor 1.22 is the only memento of the round geometry. This sets the ultimate limit on resolving detail. Two distant point sources (two stars, two cells under a microscope) each make their own Airy disk; if the disks overlap too much, they smear into one. The standard line is **Rayleigh's criterion**: two sources are *just resolved* when the center of one diffraction pattern falls on the first minimum of the other (Figure 34.4). The smallest angular separation they can have and still be told apart is therefore

Key Equation: Rayleigh's criterion

$$\theta_{\min} = \frac{\lambda}{a} \qquad (\text{slit of width } a), \tag{34.5}$$

$$\theta_{\min} = 1.22\,\frac{\lambda}{D} \qquad (\text{circular aperture of diameter } D). \tag{34.6}$$

The slit version is the easiest to picture (Figure 34.5). Light from two distant sources reaches the slit from directions an angle θ apart, and each source makes its own single-slit pattern on the screen, shifted from the other by that angle. The two sources are independent (incoherent), so their *intensities* simply add. When the shift equals λ/a, the peak of one pattern lands exactly on the first zero of the other, and the total shows two humps with a dip to about 81% between them: just resolved. Bring the sources closer and the dip fills in until only a single hump remains, as in Figure 34.4(c). A circular aperture works the same way, with its first zero at $1.22\lambda/D$ in place of λ/a.

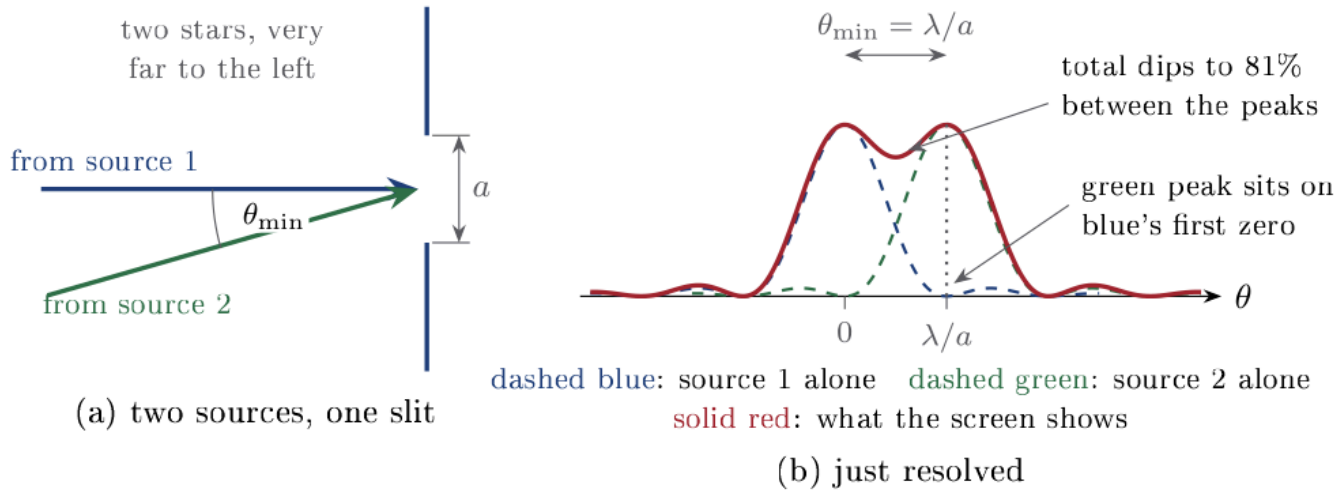


Figure 34.5. Resolving two sources with a slit. (a) Light from two distant sources reaches a slit of width a from directions $\theta_{\min}$ apart. (b) Each source alone gives a single-slit pattern (dashed); the screen shows their sum (solid). When $\theta_{\min} = \lambda/a$, the peak of one pattern sits on the first zero of the other and the total dips to about 81% between the peaks: the two sources are just resolved.

•Figure description

Resolving two sources with a single slit, in two panels. Panel a: a slit of width a, with light from two very distant stars arriving from the left: a blue ray from source 1 straight on, and a green ray from source 2 arriving at a small angle theta min to it. Panel b: intensity against angle theta. A dashed blue curve is source 1's single-slit pattern, peaked at zero. A dashed green curve is source 2's identical pattern, shifted so its peak sits at lambda over a, exactly on source 1's first zero, marked with a dotted line. A solid red curve, their sum and what the screen actually shows, has two peaks with a dip to about 81 percent between them. A double arrow above marks the peak separation theta min equals lambda over a. The panel is labeled just resolved.

The lesson is built into the formula: to resolve finer detail, use a *shorter* wavelength or a *bigger* aperture. This single relation explains why research telescopes are enormous, why electron microscopes (tiny wavelength) outresolve light microscopes, and why your pupil's few millimeters set a limit on the fine print you can read.

Example 34.2: The resolution of the eye

Light of wavelength $500\ \text{nm}$ enters an eye with a daytime pupil diameter of about $2\ \text{mm}$. (a) Estimate the eye's diffraction-limited angular resolution. (b) At a reading distance of $L = 25\ \text{cm}$, what is the smallest separation between two dots the eye can just distinguish?

Solution

(a) By Rayleigh's criterion for a circular aperture,

$$\theta_{\min} = 1.22\,\frac{\lambda}{D} = 1.22\,\frac{500\times10^{-9}\,\text{m}}{2\times10^{-3}\,\text{m}} = \boxed{3.05\times10^{-4}\,\text{rad}}, \tag{34.2.1}$$

about one arcminute: the rule of thumb for normal eyesight.

(b) The smallest resolvable separation is

$$d = L\,\theta_{\min} = (0.25\,\text{m})(3.05\times10^{-4}) = \boxed{7.6\times10^{-5}\,\text{m} \approx 0.08\,\text{mm}}. \tag{34.2.2}$$

Finer than this and the two dots blur into one, which is exactly the principle a printer relies on, and the reason a screen's pixels vanish at normal viewing distance.

Example 34.3: Telescopes, optical and radio

(a) Each Keck telescope has an effective diameter of $10\ \mathrm{m}$. Find its resolution at $600\ \mathrm{nm}$. (b) The Arecibo radio dish was $305\ \mathrm{m}$ across and observed at $\lambda = 0.75\ \mathrm{m}$. Compare its angular resolution with Keck's.

Solution

(a) For Keck,

$$\theta_{\min} = 1.22\,\frac{600\times10^{-9}\,\mathrm{m}}{10\,\mathrm{m}} = \boxed{7.3\times10^{-8}\,\mathrm{rad}}, \tag{34.2.3}$$

about fifteen thousandths of an arcsecond: extraordinarily sharp.

(b) For Arecibo,

$$\theta_{\min} = 1.22\,\frac{0.75\,\mathrm{m}}{305\,\mathrm{m}} = \boxed{3.0\times10^{-3}\,\mathrm{rad}}. \tag{34.2.4}$$

Despite its vastly larger dish, Arecibo's resolution is about $40{,}000$ times *worse* than Keck's, because its wavelength is larger by an even greater factor. Resolution is set by the *ratio* λ/D, not by size alone: a humbling fact for radio astronomy, which compensates by linking dishes continents apart into a single interferometer.

Checkpoint 34.2

Two headlights of a distant car are $1.5\ \mathrm{m}$ apart. Using the eye's resolution from the worked example, estimate the greatest distance at which you could still see them as two separate lights rather than one. (Set the angular separation equal to $\theta_{\min}$.)

Answer

Answers to Checkpoints

Checkpoint 34.2. The headlights are just resolved when their angular separation equals $\theta_{\min} = 3.05\times10^{-4}\ \mathrm{rad}$. For small angles the separation is s/L, so $L = s/\theta_{\min} = (1.5\ \mathrm{m})/(3.05\times10^{-4}) \approx 4.9\times10^{3}\ \mathrm{m}$, about $5\ \mathrm{km}$. Beyond that, the two lights blur into one.

This page titled 34.2: Diffraction by a Circular Aperture is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.3: Diffraction by a Double Slit

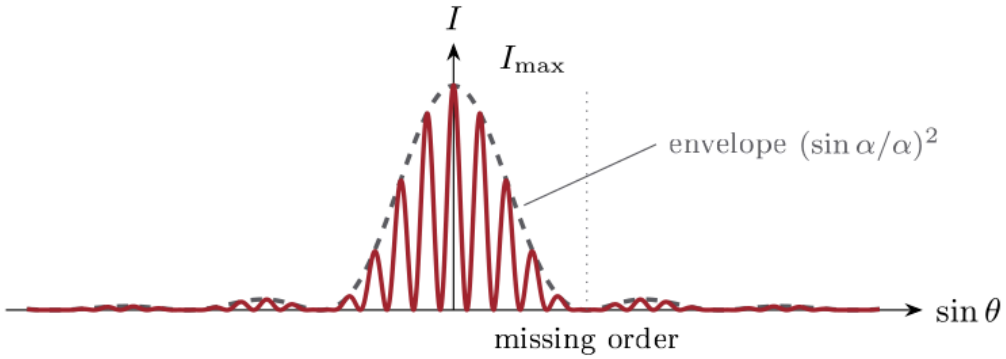


Figure 34.6. Double-slit diffraction (here $d/a = 5$). Fine interference fringes ($\cos^2\beta$) ride under the single-slit envelope $(\sin\alpha/\alpha)^2$ (dashed). Where the envelope hits zero, an interference fringe is suppressed: a "missing order."

•Figure description

A graph of intensity against sine theta for a real double slit with slit separation five times the slit width. A dashed gray curve is the broad single-slit diffraction envelope, sine alpha over alpha squared, peaking at the center and falling to zero on either side. Beneath it a red curve shows the actual pattern: rapid cosine-squared interference fringes whose peaks all touch the envelope, so the fringes are tall near the center and progressively dimmer outward. At the point where the envelope reaches zero, a fringe that would otherwise be bright is entirely suppressed, marked missing.

In Chapter 33 we treated the two slits as if each were a mathematical point, and found evenly spaced interference fringes of equal height. Real slits have width, and each one diffracts. The true double-slit pattern is therefore *both* effects at once: the closely spaced interference fringes (set by the slit *separation* d) ride underneath the broad single-slit diffraction envelope (set by the slit *width* a). The intensity is the product of the two patterns we have already met:

$$I(\theta) = I_{\max}\ \underbrace{\cos^2\beta}_{\text{interference}}\ \underbrace{\left(\frac{\sin\alpha}{\alpha}\right)^2}_{\text{diffraction}}, \qquad \beta = \frac{\pi d}{\lambda}\sin\theta, \quad \alpha = \frac{\pi a}{\lambda}\sin\theta. \tag{34.7}$$

The rapid $\cos^2\beta$ factor makes the fine interference fringes; the slow $(\sin\alpha/\alpha)^2$ factor is an envelope that dims the outer fringes (Figure 34.6).

A striking consequence is the **missing order**. Wherever a diffraction *minimum* ($a\sin\theta = m_a\lambda$) coincides with an interference *maximum* ($d\sin\theta = m_d\lambda$), the envelope zeroes out a fringe that "should" be bright. Dividing the two conditions, this happens when $m_d/m_a = d/a$. If d/a is a whole number, every (d/a)-th interference fringe vanishes.

Example 34.4: Counting fringes under the envelope

A double slit has slit separation $d = 0.100\ \text{mm}$ and slit width $a = 0.0200\ \text{mm}$. How many bright interference fringes fall within the central diffraction peak?

Solution

The central diffraction peak ends at the first diffraction minimum, $a\sin\theta = \lambda$. The interference fringe at order m_d sits at $d\sin\theta = m_d\lambda$. The last bright fringe inside the envelope is the largest m_d with $d\sin\theta < \lambda(d/a)$, i.e. $m_d < d/a$:

$$\frac{d}{a} = \frac{0.100\,\text{mm}}{0.0200\,\text{mm}} = 5. \tag{34.3.1}$$

So the $m_d = \pm5$ fringes land exactly on the envelope's first zeros and are *missing*; the bright fringes that survive inside the central peak are $m_d = 0, \pm1, \pm2, \pm3, \pm4$, that is, $\boxed{9\ \text{bright fringes}}$. This is the quantitative version of the chapter's opening question: the two-slit fringes are still there, but the single-slit diffraction of each opening decides how many you actually see.

This page titled 34.3: Diffraction by a Double Slit is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.4: Diffraction Gratings

Now answer the rest of the opening question: not two slits, but thousands. A **diffraction grating** is a large number N of identical, equally spaced slits, ruled at a spacing d (often quoted as grooves per millimeter). The condition for a bright line is the same as for two slits; all the slits are in phase when the wave from each is a whole number of wavelengths ahead of its neighbor's (Figure 34.7):

Key Equation: Grating equation

$$d\sin\theta = m\lambda, \qquad m = 0, \pm1, \pm2, \ldots \tag{34.8}$$

What changes with many slits is the *sharpness*. With two slits the brightness varies as a gentle $\cos^2$; with thousands, the maxima become razor-thin lines separated by broad darkness, because away from an exact maximum the N wavelets fan out over all phases and cancel almost completely. The width of each line narrows as the number of illuminated slits grows; its angular half-width is

$$\Delta\theta_{\text{hw}} = \frac{\lambda}{Nd\cos\theta}. \tag{34.9}$$

The more lines the light crosses, the sharper the maxima, which is what makes a grating a precision tool for measuring wavelength.

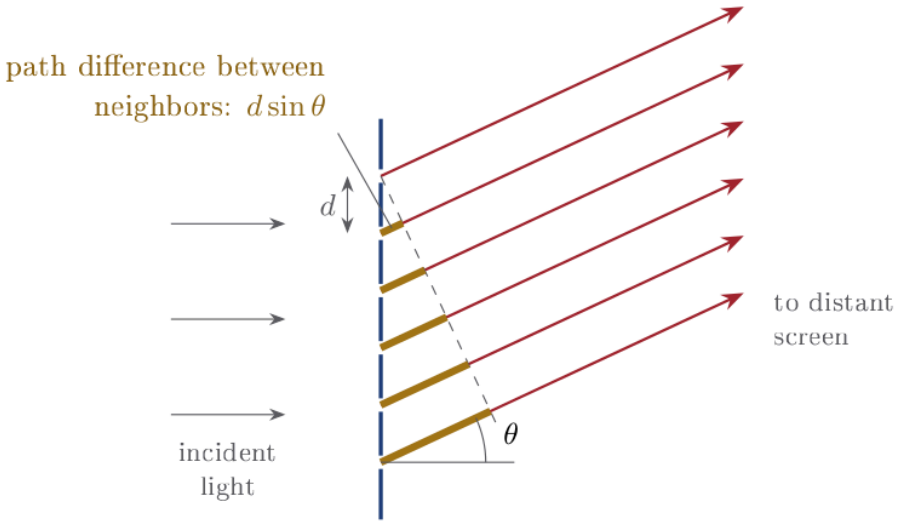


Figure 34.7. A diffraction grating. Light leaving neighboring slits at angle θ differs in path by $d\sin\theta$ (gold), so slit k is $k\ d\sin\theta$ behind the top one. When $d\sin\theta = m\lambda$, *all* N waves arrive in step and add to a bright line; at any other angle their phases spread around the circle and nearly cancel, which is why the lines are so sharp.

•Figure description

A diffraction grating seen edge-on: a vertical barrier pierced by six equally spaced narrow slits, a distance d apart, with plane light arriving from the left. From every slit a red ray leaves at the same angle theta above the horizontal, heading toward a distant screen. A dashed line drawn from the top slit perpendicular to the rays crosses each lower ray a little farther out, and the stretch of each ray between its slit and that line is highlighted in gold: one unit of d sine theta for the second slit, two for the third, and so on. When d sine theta is a whole number of wavelengths, every slit's wave arrives in step with all the others and a bright line forms.

That precision is the basis of **spectroscopy**. Because the angle of each bright line depends on wavelength, a grating fans a beam of light out into its spectrum, spreading the colors far more cleanly than a prism. Every kind of atom emits and absorbs its own fixed set of wavelengths (a barcode set by its electron energy levels), so the pattern of lines in starlight, sunlight, or a flame is a fingerprint of the elements present. It is by grating spectroscopy that we know the composition of the Sun, the planets, and galaxies we will never visit: we read it directly off their light.

Example 34.5: A helium–neon laser on a grating

Light from a helium–neon laser ($\lambda = 632.8\ \text{nm}$) strikes a grating with 6000 grooves per centimeter at normal incidence. (a) Find the angles of the first- and second-order maxima. (b) Is there a third order?

Solution

The slit spacing is the reciprocal of the groove density:

$$d = \frac{1\,\text{cm}}{6000} = 1.667 \times 10^{-4}\,\text{cm} = 1.667 \times 10^{-6}\,\text{m}. \tag{34.4.1}$$

(a) From $d\sin\theta = m\lambda$, $\sin\theta = m\lambda/d$. For $m = 1$,

$$\sin\theta_1 = \frac{632.8 \times 10^{-9}}{1.667 \times 10^{-6}} = 0.3797, \quad \theta_1 = \boxed{22.3^\circ}. \tag{34.4.2}$$

For $m = 2$, $\sin\theta_2 = 0.7594$, so $\theta_2 = \boxed{49.4^\circ}$.

(b) For $m = 3$, $\sin\theta_3 = 3(0.3797) = 1.139 > 1$, which is impossible. $\boxed{\text{There is no third order}}$: the grating cannot bend light past 90°, so only a finite number of orders exist.

Checkpoint 34.3

White light (roughly 400 to $700\ \text{nm}$) falls on a grating. In the first-order spectrum, which color is diffracted to the *largest* angle: red or violet? Use the grating equation to justify your answer, and contrast it with how a glass prism orders the colors.

Answer

Answers to Checkpoints

Checkpoint 34.3. Red. The grating equation for first order, $d\sin\theta = \lambda$, gives $\sin\theta = \lambda/d$, so the longest wavelength is diffracted to the largest angle: violet lands closest to the central maximum and red farthest out. A prism orders the colors the other way around: it separates colors by refraction, and glass's index is larger for violet than for red, so violet is bent the most.

This page titled 34.4: Diffraction Gratings is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.5: Diffraction of X-Rays by Crystals

A grating works only when its spacing is comparable to the wavelength of the light. Visible light has $\lambda \sim 500\ \text{nm}$, and the finest ruled gratings reach spacings near that. But the atoms in a solid are spaced about 0.1 to $0.3\ \text{nm}$ apart: a thousand times finer. No ruling engine can match that; yet nature has already built the grating for us. A crystal is a regular, three-dimensional lattice of atoms, and if we illuminate it with radiation of comparably tiny wavelength (**X-rays**, $\lambda \sim 0.1\ \text{nm}$) the planes of atoms act as a diffraction grating in depth.

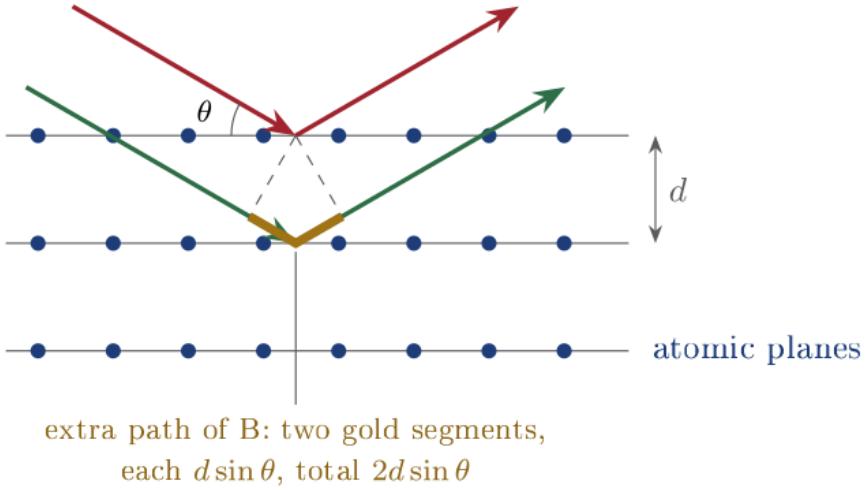


Figure 34.8. Bragg diffraction. X-rays reflecting from successive atomic planes (spacing d) reinforce when the extra path $2d\sin\theta$ is a whole number of wavelengths. The angle θ is measured from the plane.

•Figure description

Bragg diffraction from a crystal. Three horizontal rows of blue atoms represent parallel atomic planes, a distance d apart. A red X-ray comes down from the upper left, reflects off the top plane, and leaves up to the upper right at the same angle theta, measured from the plane itself rather than from the normal. A green X-ray, parallel to the first, passes the top plane and reflects off the second plane instead, then leaves parallel to the red one. The green ray has traveled an extra distance of 2 d sine theta, and the two reflections reinforce only when that extra path is a whole number of wavelengths.

Picture the atoms arranged in parallel planes a distance d apart (Figure 34.8). X-rays reflect weakly from each plane. Rays reflecting from two adjacent planes travel different distances: the lower ray goes an extra $2d\sin\theta$, where θ is measured from the *plane* (not from the normal, as in optics; this is the crystallographers' convention). The reflections from all the planes reinforce only when this extra path is a whole number of wavelengths:

Key Equation: Bragg's law

$$2d\sin\theta = m\lambda, \qquad m = 1, 2, 3, \ldots \tag{34.10}$$

Run the logic forward and Bragg's law tells you where the diffracted beams go for a known crystal: the basis of the crystal spectrometers that sort X-rays by wavelength. Run it backward and it becomes the most powerful structural tool in science: measure the angles at which a crystal diffracts X-rays of known wavelength, and you deduce the spacing and arrangement of its atoms. This is how the structure of DNA was read from a diffraction photograph, and how the architecture of countless proteins, minerals, and materials has been mapped since.

Example 34.6: Spacing of atomic planes

A beam of X-rays of wavelength $0.150\ \text{nm}$ produces a first-order Bragg reflection from a crystal at an angle of 15.4° from the atomic planes. Find the spacing of the planes.

Solution

Solve Bragg's law, equation (34.10), for d with $m = 1$:

$$d = \frac{m\lambda}{2\sin\theta} = \frac{(1)(0.150\,\text{nm})}{2\sin 15.4^\circ} = \frac{0.150\,\text{nm}}{2(0.2656)} = \boxed{0.282\,\text{nm}}. \tag{34.5.1}$$

This is the spacing in rock salt (NaCl): about the size of an atom, far too small for any optical instrument to resolve, yet read off cleanly from the diffraction angle.

This page titled 34.5: Diffraction of X-Rays by Crystals is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.S: Diffraction (Summary)

Summary

- **Single-slit diffraction**: minima at $a\sin\theta = m\lambda$ ($m = 1, 2, \ldots$; no $m = 0$). The central maximum has full width $w = 2\lambda L/a$ on a distant screen: narrower slits spread the light more.
- **Intensity**: $I = I_{\max}(\sin\alpha/\alpha)^2$ with $\alpha = (\pi a/\lambda)\sin\theta$. Almost all the light is in the central peak; secondary maxima are faint.
- **Circular aperture**: first dark ring at $\sin\theta = 1.22\lambda/D$. **Rayleigh's criterion** sets the resolution limit: $\theta_{\min} = \lambda/a$ (slit) or $1.22\lambda/D$ (circular). Finer detail needs shorter λ or larger D.
- **Double-slit diffraction**: interference fringes $\cos^2\beta$ ride under the single-slit envelope $(\sin\alpha/\alpha)^2$; $\beta = (\pi d/\lambda)\sin\theta$. Fringes vanish ("missing orders") where $m_d/m_a = d/a$.
- **Diffraction grating**: N slits at spacing d give sharp bright lines at $d\sin\theta = m\lambda$, with line half-width $\lambda/(Nd\cos\theta)$, the basis of spectroscopy and the chemical fingerprinting of light.
- **X-ray diffraction**: a crystal's atomic planes (spacing d) diffract X-rays per **Bragg's law**, $2d\sin\theta = m\lambda$ (θ from the plane), the key to atomic and molecular structure.

Key Equations

Single-slit minima

$$a\sin\theta = m\lambda, \qquad m = 1, 2, 3, \ldots \quad (\text{no } m = 0) \tag{34.S.1}$$

Width of the central maximum

$$w = \frac{2\lambda L}{a} \tag{34.S.2}$$

Single-slit intensity

$$I(\theta) = I_{\max}\left(\frac{\sin\alpha}{\alpha}\right)^2, \qquad \alpha = \frac{\pi a}{\lambda}\sin\theta \tag{34.S.3}$$

Circular-aperture first minimum

$$\sin\theta = 1.22\,\frac{\lambda}{D} \tag{34.S.4}$$

Rayleigh's criterion

$$\begin{aligned} \theta_{\min} &= \frac{\lambda}{a} && (\text{slit of width } a) \\ \theta_{\min} &= 1.22\,\frac{\lambda}{D} && (\text{circular aperture of diameter } D) \end{aligned} \tag{34.S.5}$$

Double slit with finite slit width

$$I(\theta) = I_{\max}\cos^2\beta\left(\frac{\sin\alpha}{\alpha}\right)^2, \qquad \beta = \frac{\pi d}{\lambda}\sin\theta, \quad \alpha = \frac{\pi a}{\lambda}\sin\theta \tag{34.S.6}$$

Missing orders

$$\frac{m_d}{m_a} = \frac{d}{a} \tag{34.S.7}$$

Grating equation

$$d \sin\theta = m\lambda, \qquad m = 0, \pm 1, \pm 2, \ldots \tag{34.S.8}$$

Angular half-width of a grating line

$$\Delta\theta_{\text{hw}} = \frac{\lambda}{Nd\cos\theta} \tag{34.S.9}$$

Bragg's law

$$2d \sin\theta = m\lambda \qquad (\theta \text{ measured from the atomic plane}) \tag{34.S.10}$$

This page titled 34.S: Diffraction (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

34.E: Diffraction (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Treat angles as small where appropriate ($\sin\theta \approx \tan\theta = y/L$), and take visible light to span roughly 400–$700\ \text{nm}$.

Section 34.1 — Single-Slit Diffraction

34.1 ★ Light of wavelength $633\ \text{nm}$ passes through a slit $0.085\ \text{mm}$ wide. At what angle is the first diffraction minimum?

34.2 ★★ A slit $0.040\ \text{mm}$ wide is lit by $550\ \text{nm}$ light, with a screen $1.5\ \text{m}$ away. Find the width of the central bright fringe.

34.3 ★★ *[Conceptual.]* As a single slit is made narrower and narrower, describe what happens to the diffraction pattern, and explain why in terms of the condition $a\sin\theta = m\lambda$.

34.4 ★★ A single slit $0.060\ \text{mm}$ wide is lit by a laser of unknown wavelength. On a screen $1.5\ \text{m}$ away, the central bright band is $2.4\ \text{cm}$ wide. (a) Find the wavelength of the laser. (b) The slit is replaced by one $0.040\ \text{mm}$ wide. How wide is the central bright band now?

Section 34.1.1 — Intensity in Single-Slit Diffraction

34.5 ★★ In a single-slit pattern, find the intensity (as a fraction of $I_{\max}$) at the angle where $\alpha = \pi/2$.

34.6 ★★★ *[Conceptual.]* Explain why the single-slit intensity formula $I = I_{\max}(\sin\alpha/\alpha)^2$ gives a maximum at $\alpha = 0$ even though $\sin\alpha/\alpha$ is the indeterminate form $0/0$ there. What is the value of the limit?

34.7 ★★★ In single-slit diffraction the intensity is $I = I_{\max}(\sin\alpha/\alpha)^2$, with $\alpha = (\pi a/\lambda)\sin\theta$. The first secondary maximum falls near $\alpha = 3\pi/2$. Find the ratio of its intensity to that of the central maximum.

Section 34.2 — Diffraction by a Circular Aperture

34.8 ★ A telescope has an aperture of diameter $5.0\ \text{m}$. Find its diffraction-limited angular resolution at $550\ \text{nm}$.

34.9 ★★ *[Real-world application: astronomy.]* The Hubble Space Telescope has a mirror $2.4\ \text{m}$ in diameter. (a) Find its angular resolution at $500\ \text{nm}$. (b) At the distance of the Moon ($3.8\times10^{8}\ \text{m}$), what is the smallest feature it could resolve?

34.10 ★★★ *[Real-world application: vision.]* Two stars are separated by an angle of $2.0\times10^{-5}\ \text{rad}$. What minimum aperture diameter is needed to resolve them at $550\ \text{nm}$? Could the human eye (pupil $\sim 2\ \text{mm}$) do it?

34.11 ★★ *[Real-world application: vision.]* In bright light the pupil is about $3.0\ \text{mm}$ across. Using $550\ \text{nm}$ light, find (a) the smallest angular separation the eye can resolve, and (b) the smallest separation of two dots at a reading distance of $40\ \text{cm}$.

Section 34.3 — Diffraction by a Double Slit

34.12 ★★ A double slit has slit separation $d = 0.090\ \text{mm}$ and slit width $a = 0.015\ \text{mm}$. Which interference orders are missing (suppressed by the diffraction envelope)?

34.13 ★★★ For a double slit with $d = 0.12\ \text{mm}$ and $a = 0.030\ \text{mm}$, how many bright interference fringes lie within the central diffraction maximum?

34.14 ★★★ A double slit has slit separation $0.10\ \text{mm}$ with each slit $0.020\ \text{mm}$ wide, illuminated by $500\ \text{nm}$ light. How many bright interference fringes lie within the central diffraction envelope?

Section 34.4 — Diffraction Gratings

34.15 ★ A diffraction grating has 5000 lines per centimeter. Find the angle of the first-order maximum for $600\ \text{nm}$ light.

34.16 ★★ *[Real-world application: spectroscopy.]* A grating with 8000 lines per centimeter is lit by light containing $486\ \text{nm}$ and $656\ \text{nm}$ (two hydrogen lines). Find the angular separation between these two lines in the first order.

34.17 ★★★ White light (400–$700\ \text{nm}$) falls on a grating with 4000 lines per centimeter. (a) Find the angular width of the first-order spectrum. (b) Does the first-order spectrum overlap the second-order spectrum? Justify with angles.

Section 34.5 — X-Rays & the Many-Slit Question

34.18 ★★ X-rays of wavelength $0.120\ \text{nm}$ undergo first-order Bragg reflection from a crystal whose atomic planes are spaced $0.250\ \text{nm}$ apart. At what angle (from the planes) does the reflection occur?

34.19 ★★★ *[Real-world application: materials science.]* First-order Bragg reflection of $0.165\ \text{nm}$ X-rays from a crystal is observed at 21.5° from the planes. (a) Find the interplanar spacing. (b) At what angle would the second-order reflection appear, if it exists?

34.20 ★★★ *[Conceptual — one slit to many.]* Trace the chapter's through-line. Explain in a few sentences how the pattern evolves as you go from one slit, to two slits, to a grating of thousands of slits, to a crystal's three-dimensional lattice: identifying for each what stays the same (the role of path difference) and what changes (the sharpness and dimensionality of the pattern).

Challenge Problems

34.21 ★★★ *[Real-world application: astronomy — imaging a black hole.]* The shadow of the supermassive black hole at the center of the galaxy M87 subtends about 42 microarcseconds (42×10^{-6} arcsec) as seen from Earth. Radio astronomers imaged it by combining telescopes across the planet into one instrument whose effective aperture is the separation between the dishes. Use $1\ \text{radian} = 206265$ arcsec.

a. Convert 42 microarcseconds to radians.
b. Observing at $\lambda = 1.3\ \text{mm}$, what aperture diameter does Rayleigh's criterion require to resolve a feature this small?
c. Compare your answer with Earth's diameter, $12742\ \text{km}$. Comment on what this says about how much margin the observation had.
d. A later campaign by the same collaboration observed at $0.87\ \text{mm}$ instead. Keeping the baseline at Earth's diameter, find the angular resolution at each wavelength, in microarcseconds, and explain the motivation for the shorter wavelength. What is the fundamental obstacle to improving the resolution further without leaving the planet?

34.22 ★★★ *[Real-world application: spectroscopy — resolving the sodium doublet.]* The yellow light of a sodium lamp is not one line but two, at $589.00\ \text{nm}$ and $589.59\ \text{nm}$. A grating can separate them only if its *resolving power* $R = \lambda/\Delta\lambda$ is large enough. For a grating, $R = Nm$, where N is the number of slits illuminated and m is the order.

a. Justify $R = Nm$ from the angular half-width $\Delta\theta_{\text{hw}} = \lambda/(Nd\cos\theta)$ given in this chapter: two wavelengths are just resolved when the maximum of one falls at the first minimum of the other. *(Hint: differentiate the grating equation to find how far a maximum shifts when λ changes by $\Delta\lambda$, and set that shift equal to $\Delta\theta_{\text{hw}}$.)*
b. Find the resolving power needed for the sodium doublet.
c. A grating is ruled with 600 lines per millimeter. In first order, how wide must the illuminated patch of grating be? At what angle does the doublet appear?
d. Repeat part (c) for second order. Which order would you use, and what is the practical cost of going to higher order?

34.23 ★★★ *[Real-world application: photography — the diffraction-limited aperture.]* A camera lens's f-number is $N_f = f/D$, where D is the diameter of the aperture. The lens focuses light onto a sensor a distance $\approx f$ behind it, so the diameter of the Airy disk formed there is $d = 2f\theta_{\text{min}}$, with $\theta_{\text{min}} = 1.22\lambda/D$. Take $\lambda = 550\ \text{nm}$.

a. Show that $d = 2.44\ \lambda\ N_f$: the focal length drops out entirely, so the diffraction spot depends only on the f-number.
b. Evaluate d at $f/2.8$ and at $f/16$.
c. A full-frame sensor $36\ \text{mm}$ wide with 8400 pixels across has a pixel pitch of about $4.3\ \mu\text{m}$. Find the f-number at which the

Airy disk first spans two pixels, so that stopping down further can no longer add detail.

d. Landscape photographers nevertheless routinely shoot at $f/16$. Explain what they are buying with the resolution they give up, and why the trade is usually worth it.

This page titled 34.E: Diffraction (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35: RELATIVITY

Einstein and Newton are heading out for the evening. Einstein phones to say he will wait, at rest, in his car until Newton arrives so they can drive to the bar together. Newton points out that he doesn't drive. "No matter," says Einstein. Ten minutes later Einstein pulls up at Newton's house, rolls down the window, and says, "Thanks for coming to pick me up!" Newton, baffled, protests: he never moved; Einstein clearly drove over. Einstein only smiles and says: *prove it*.

It sounds like a joke, but the challenge is serious, and answering it overturned three centuries of physics. How *would* you prove which of two people moving at constant velocity is "really" moving? Every test you can think of fails. The bumps of the road would betray motion, but those come from acceleration, so let us agree to compare only **inertial frames**, frames moving at constant velocity (this restriction is what makes the theory *special* relativity). Within that rule, toss an apple, launch a toy rocket, pour a drink: every mechanical experiment comes out the same whether you call yourself moving or at rest, because every particle carries inertia. Newton himself knew this. The trouble began when physicists asked the same question of *light*, and the answer broke the old physics. This chapter follows that question to its conclusions: that moving clocks run slow, moving objects shorten, simultaneity dissolves, mass and energy are the same thing, and gravity is the bending of spacetime itself.

Learning Objectives

- Explain the experimental and conceptual roots of special relativity, including the Michelson–Morley result.
- State Einstein's two postulates and the meaning of an inertial frame.
- Apply time dilation and length contraction using the Lorentz factor γ.
- Identify which frame measures the proper time and which measures the proper length.
- Add velocities relativistically and transform coordinates with the Lorentz equations.
- Explain the relativity of simultaneity.
- Apply the relativistic Doppler effect to light, including cosmological redshift.
- Use relativistic momentum and the mass–energy relation $E = \gamma mc^2$.
- Describe the equivalence principle and the basic predictions of general relativity.

We will return to Einstein's challenge ("prove it") throughout, for it is the thread that ties the whole theory together.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

35.1: History and Motivation for Special Relativity
35.2: Consequences of Special Relativity
35.3: Lorentz Transformations
35.4: Doppler Effect for Light
35.5: Momentum and Energy
35.6: General Relativity
35.S: Relativity (Summary)
35.E: Relativity (Problems)

This page titled 35: Relativity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.1: History and Motivation for Special Relativity

Return to the challenge. Mechanical experiments cannot reveal absolute motion; this is the old **principle of relativity**, known since Galileo: the laws of mechanics are the same in every inertial frame. But what about *waves*? A wave's speed is set by its medium, not its source, so perhaps motion through the medium is detectable. Shout while driving and the sound still leaves your mouth at the speed of sound *relative to the air*; the air is the referee. If you could find the medium and measure your motion through it, you could finally answer Einstein.

Light seemed the perfect probe. Maxwell's equations (Chapter 27) predict that light is an electromagnetic wave traveling at a definite speed c, and a wave, surely, needs a medium. Physicists named the supposed medium the **luminiferous aether** and set out to measure Earth's motion through it. The idea was simple: if Earth plows through the aether, light sent "downstream" should travel faster than light sent "across," just as a swimmer is faster downstream than across a river. In 1887 Michelson and Morley built an exquisitely sensitive interferometer (Chapter 33) to detect the difference.

It found nothing. No matter the direction or the season, the fringes refused to shift: light traveled at the same speed c in every direction. The aether wind simply was not there. This null result (one of the most consequential failures in the history of science) left two facts that classical physics could not hold together at once: that constant-velocity motion is undetectable, and that light always travels at c. Einstein's genius in 1905 was to stop treating the second fact as a problem and start treating it as a *law*.

Einstein's Principle of Relativity

Einstein resolved the conflict by elevating both facts to postulates and following them without flinching:

Key Equation: Einstein's two postulates (1905)

1. **The principle of relativity:** the laws of physics are the same in all inertial reference frames.
2. **The constancy of the speed of light:** the speed of light in vacuum has the same value c in all inertial frames, independent of the motion of the source or observer.

That is the entire theory. There is no third assumption. The first postulate is Galileo's, extended from mechanics to *all* of physics, so that no experiment of any kind can detect absolute motion: answering Einstein's challenge with a flat "you can't." The second is the Michelson–Morley result taken at face value: everyone, however they move, measures the same c. The postulates look modest, even bland. But holding both at once forces us to give up something we never thought to question (the idea that time and space are the same for everyone) and the consequences are profound.

This page titled 35.1: History and Motivation for Special Relativity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.2: Consequences of Special Relativity

Here is the crux. Speed is distance over time. If two observers in relative motion must both measure the *same* speed c for the same light beam (even though one of them is chasing it) then they cannot agree on the distances and times that make up that speed. Something has to give, and what gives is our intuition that space and time are absolute. The second postulate forces three intertwined consequences, which the rest of this chapter unpacks: moving clocks run slow (*time dilation*), moving objects shorten along their motion (*length contraction*), and events that are simultaneous for one observer are not for another (*the relativity of simultaneity*).

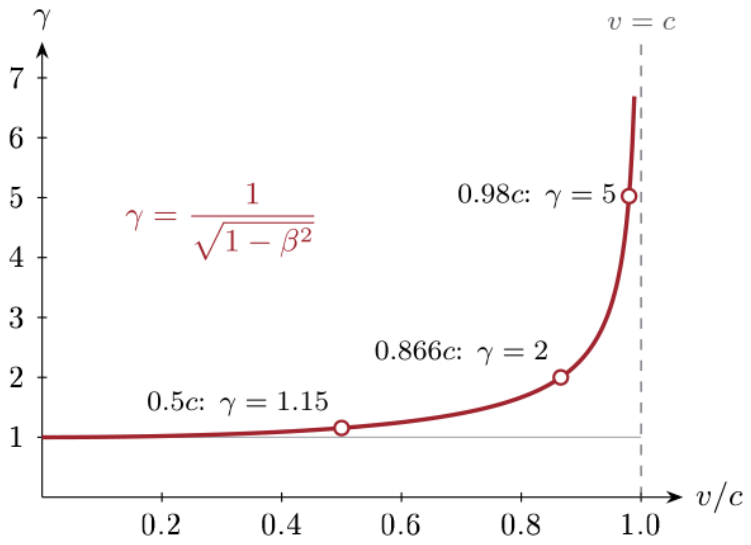


Figure 35.1. The Lorentz factor γ versus speed. It stays near 1 until v is a large fraction of c, then diverges as $v \to c$. Relativistic effects are negligible at everyday speeds and overwhelming near light speed.

•Figure description

A graph of the Lorentz factor gamma against speed as a fraction of the speed of light, v over c, running from 0 to 1, with gamma from 0 to 7 on the vertical axis. The curve starts at gamma equals 1 and stays almost flat and indistinguishable from 1 out to about v equals 0.5 c, where a marked point reads gamma equals 1.15. Beyond that it bends upward: marked points show gamma equals 2 at 0.866 c and gamma equals 5 at 0.98 c. The curve then climbs almost vertically, passing gamma equals 7 at 0.99 c, alongside a dashed vertical asymptote at v equals c, where it diverges. Relativistic effects are invisible at everyday speeds and overwhelming only in the last sliver before light speed.

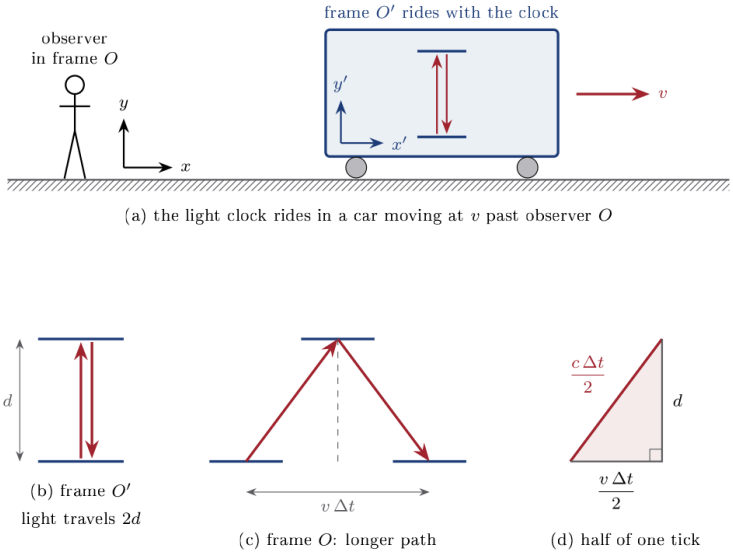
Every one of these effects is governed by a single dimensionless quantity, the **Lorentz factor**. Writing $\beta = v/c$ for speed as a fraction of light speed,

Key Equation: Lorentz factor

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1 - \beta^2}}. \tag{35.1}$$

Because $v < c$ always, γ is always greater than or equal to 1. At everyday speeds it is so close to 1 that relativity is invisible (Figure 35.1); only as v approaches c does γ climb steeply toward infinity. This is why the world looks Newtonian: we crawl. At 10% of light speed γ is just 1.005; the effects become dramatic only in the last sliver of the speed range.

Time Dilation



(a) the light clock rides in a car moving at v past observer O

(b) frame O' light travels $2d$

(c) frame O: longer path

(d) half of one tick

Figure 35.2. The light clock. (a) The clock rides in a car, frame O', that moves at speed v past an observer standing on the ground, frame O. (b) In the clock's own frame O' the pulse bounces straight up and down a total distance $2d$ in proper time Δt_p. (c) In frame O the mirrors move during the tick, so the pulse takes a longer diagonal path; since its speed is still c, the tick takes longer: time dilation. (d) Half of one tick is a right triangle with hypotenuse $c\ \Delta t/2$ and legs d and $v\ \Delta t/2$.

•Figure description

The light clock, in four panels. Panel a, the set-up: an observer stands on the ground in frame O, with x and y axes beside them. To the right a car rolls along the ground at speed v; inside it is a light clock, two horizontal mirrors with a red pulse bouncing straight up and down, and the car carries its own x prime and y prime axes, labeled frame O prime rides with the clock. Panel b, frame O prime: two mirrors a distance d apart with the pulse going straight up and back, so light travels 2 d. Panel c, frame O: the lower mirror is drawn at its start and end positions, a distance v times delta t apart, with the upper mirror midway; the pulse travels up a diagonal to the upper mirror and down another, symmetric diagonal, a longer path, and a dashed vertical line drops from the top of the path. Panel d, half of one tick: a right triangle with vertical leg d, horizontal leg v delta t over 2, and hypotenuse c delta t over 2, with a right-angle mark where the two legs meet.

To see time stretch, build the simplest possible clock: a **light clock**, in which a pulse bounces between two mirrors a distance d apart, ticking once per round trip (Figure 35.2). Ride alongside it and the pulse goes straight up and back, covering $2d$; the time for one tick is $\Delta t_p = 2d/c$. We call this the **proper time** Δt_p: the time between two events (here, two ticks) measured in the frame where they happen at the *same place*.

Now watch the same clock fly past you at speed v: the clock rides in frame O' while you stand in frame O (Figure 35.2(a)). The pulse still moves at c (that is the unbreakable postulate), but now it must travel a longer, diagonal path, because the mirrors slide sideways between bounces (Figure 35.2(c)). Longer path at the same speed means a longer tick.

To find out how much longer, call the time for one tick in your frame Δt. The first half of the tick carries the pulse from the bottom mirror to the top one. In that time, $\Delta t/2$, the pulse travels a distance $c\ \Delta t/2$ along the diagonal while the clock slides $v\ \Delta t/2$ sideways. The mirror separation d is the same in both frames, because it is perpendicular to the motion (Section 35.2 explains why). These three distances form the right triangle of Figure 35.2(d), so the Pythagorean theorem gives

$$\left(\frac{c\,\Delta t}{2}\right)^2 = d^2 + \left(\frac{v\,\Delta t}{2}\right)^2 . \tag{35.2.1}$$

Collect the Δt terms and solve:

$$\left(\frac{\Delta t}{2}\right)^2 (c^2 - v^2) = d^2$$
$$\Delta t = \frac{2d}{\sqrt{c^2 - v^2}} = \frac{2d/c}{\sqrt{1 - v^2/c^2}}. \tag{35.2.2}$$

The numerator, $2d/c$, is exactly the tick Δt_p measured by someone riding with the clock, and the denominator is $1/\gamma$. That is the central result:

Key Equation: Time dilation

$$\Delta t = \gamma\,\Delta t_p = \frac{\Delta t_p}{\sqrt{1 - v^2/c^2}}. \tag{35.2}$$

Since $\gamma \geq 1$, the moving clock's ticks are *spread out* as seen by you: a moving clock runs slow. The proper time Δt_p is always the shortest: the time read by the single clock present at both events.

This is not a quirk of light clocks; *all* clocks, including biological ones, must agree, or you could tell which clock was "really" moving and violate the first postulate. And it is measured every day. Muons created high in the atmosphere live only about 2.2 μs at rest: far too short, at near-light speed, to reach the ground. Yet they arrive in abundance, because their internal clocks run slow in our frame and they live long enough to make the trip. Atomic clocks have been flown around the world on airliners and carried aboard the International Space Station, and they come back out of step with their twins on the ground by just the amounts relativity predicts.

Example 35.1: A speeding pendulum

A pendulum's period is 3.00 s in its own rest frame. What period does an observer measure who moves at $0.960c$ relative to the pendulum?

Solution

The 3.00 s is the proper time (measured where the pendulum sits). First find γ at $\beta = 0.960$:

$$\gamma = \frac{1}{\sqrt{1 - 0.960^2}} = \frac{1}{\sqrt{1 - 0.9216}} = \frac{1}{\sqrt{0.0784}} = 3.57. \tag{35.2.3}$$

Then

$$\Delta t = \gamma\,\Delta t_p = (3.57)(3.00\,\text{s}) = \boxed{10.7\,\text{s}}. \tag{35.2.4}$$

The moving observer sees the pendulum swing in slow motion: more than three times slower.

Example 35.2: The late employee's excuse

You drive to a meeting at 30 m/s. Your clock reads a 5.0-hour trip, but you arrive late and blame time dilation: your car clock, you claim, ran slow. How much time really passed on your boss's Earth-bound clock?

Solution

At $v = 30$ m/s, $\beta = v/c = 1.0 \times 10^{-7}$, so γ differs from 1 by only about $\frac{1}{2}\beta^2 = 5 \times 10^{-15}$. The boss's clock reads

$$\Delta t = \gamma\,\Delta t_p, \qquad \Delta t - \Delta t_p = (\gamma - 1)\Delta t_p \approx (5 \times 10^{-15})(5.0 \times 3600\,\text{s}) \approx \boxed{9 \times 10^{-11}\,\text{s}}. \tag{35.2.5}$$

Your excuse buys you about 90 picoseconds. Time dilation is utterly real but utterly negligible at highway speeds, which is exactly why no one noticed it for all of human history.

Checkpoint 35.1

Astronauts are paid by the elapsed time of a long, near-light-speed voyage. To be paid *more*, should the crew insist on being paid according to an Earth-based clock or their own ship's clock? Explain using which clock measures the proper time of the journey.

Answer

Length Contraction

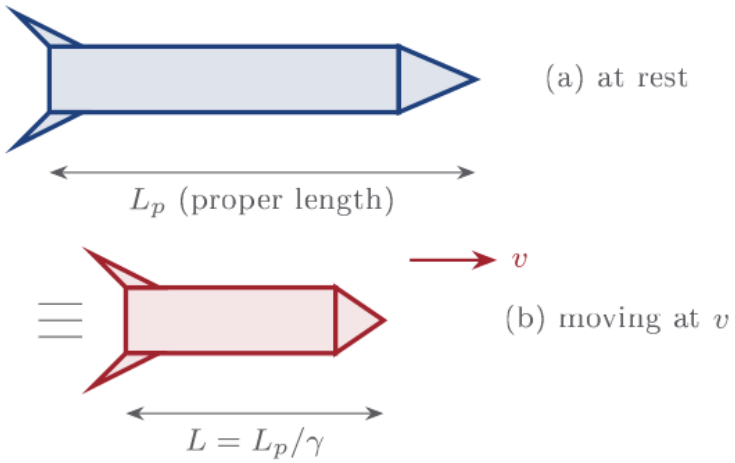


Figure 35.3. Length contraction. A spaceship of proper length L_p in its own rest frame (a) is measured shorter, $L = L_p/\gamma$, when it moves at speed v along its length (b). Only the dimension along the motion shrinks; the ship is no thinner.

•Figure description

Two drawings of the same spaceship. In panel (a), at rest, the ship is drawn in blue with its full length marked beneath as L sub p, the proper length. In panel (b), moving at speed v to the right, indicated by an arrow and by motion lines behind it, the identical ship is drawn in red and is visibly shorter nose to tail, its length marked as L equals L sub p over gamma. The two ships are exactly the same height and the fins are the same size -- only the dimension along the motion has shrunk.

If observers disagree about time, they must also disagree about length, because length is measured with clocks and light. An object's **proper length** L_p is its length measured in its own rest frame. Measured from a frame in which the object moves at speed v *along its length*, it is shorter (Figure 35.3):

Key Equation: Length contraction

$$L = \frac{L_p}{\gamma} = L_p\sqrt{1 - v^2/c^2}. \tag{35.3}$$

Only the dimension *along* the motion contracts; heights and widths are unchanged. So a spaceship streaking past looks shortened nose-to-tail but no thinner, and a meter stick flown lengthwise past you measures less than a meter, though to someone riding with it, it is a perfectly ordinary meter stick, and it is *you* who looks contracted.

Where length contraction comes from. Length contraction is not a new assumption; it follows directly from time dilation. Picture a trip from Earth to a star a distance L_p away, measured in the frame where Earth and star are at rest (Figure 35.4).

- *Earth's frame.* The ship moves at v and arrives after $\Delta t = L_p/v$ on Earth's clocks.
- *The ship's clock.* The two events, leaving Earth and reaching the star, both happen right at the ship, so a single ship clock is present at both. That clock reads the proper time, $\Delta t_p = \Delta t/\gamma$.
- *The ship's frame.* The ship sits still while Earth and the star stream past at speed v. The star arrives a time Δt_p after Earth departs, so in this frame the distance between them is

$$L = v\,\Delta t_p = \frac{v\,\Delta t}{\gamma} = \frac{L_p}{\gamma}. \tag{35.2.6}$$

Both observers agree on the relative speed v and on what the ship's clock reads at arrival. The only way for the arithmetic to work is for the ship to measure a shorter distance.

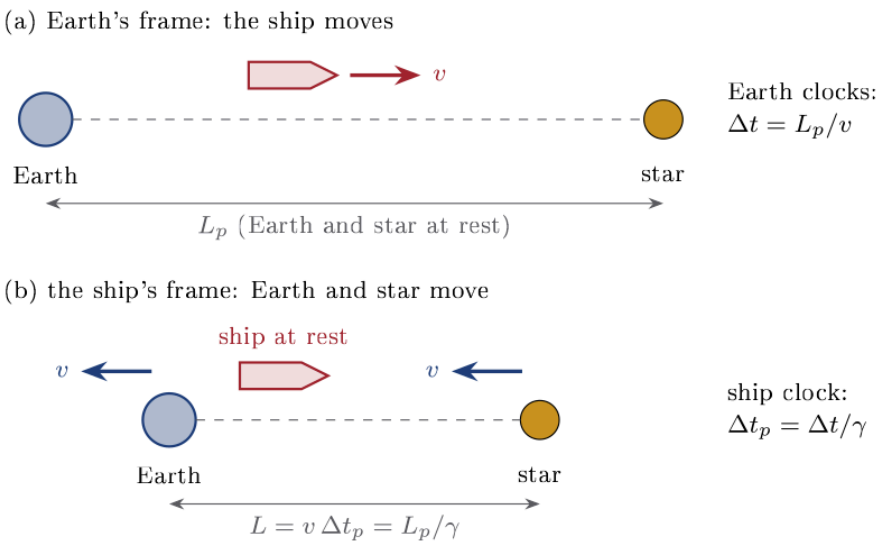


Figure 35.4. Length contraction from time dilation. (a) In Earth's frame, Earth and the star are at rest a distance L_p apart, and the trip takes $\Delta t = L_p/v$. (b) In the ship's frame, Earth and the star move past at v. The ship's clock reads only the proper time $\Delta t_p = \Delta t/\gamma$ for the trip, so the distance it measures is $L = v\ \Delta t_p = L_p/\gamma$.

•Figure description

Length contraction derived from time dilation, in two panels. Panel a, Earth's frame: Earth on the left and a star on the right, at rest a distance L sub p apart, joined by a dashed line, with a small red ship partway between them moving right at speed v. A note reads Earth clocks: delta t equals L sub p over v. Panel b, the ship's frame: the red ship is at rest, while Earth and the star both move left at speed v, shown by blue arrows. The distance between Earth and star is now visibly shorter, labeled L equals v delta t sub p equals L sub p over gamma. A note reads ship clock: delta t sub p equals delta t over gamma.

Measuring a length always involves clocks. To measure a *moving* object, you must mark where its front and back are *at the same instant*, and "the same instant" at two different places is exactly what observers in relative motion disagree about (Section 35.3). Each observer marks the two ends at moments the other considers different, so they get different lengths. Lengths *perpendicular* to the motion are immune. Two observers can hold meter sticks upright and let them pass: the tips meet at one place and one moment, so both must agree on whether the tips line up. If a moving stick shrank sideways, each observer would see the *other's* stick as shorter, a contradiction. So neither shrinks, and that is why the mirror separation d in the light clock is the same in both frames.

Example 35.3: A trip to Sirius

Sirius is about 8.6 light-years from Earth; to keep the arithmetic simple, take the distance as 8.0 light-years. An astronaut travels there at $0.80c$ and measures the one-way trip to take only 6.0 years. How is a 6.0-year trip reconciled with an 8.0-light-year distance?

Solution

At $\beta = 0.80$, $\gamma = 1/\sqrt{1-0.64} = 1/0.60 = 1.67$.

From Earth's frame: the distance is $8.0\ \text{ly}$ and the ship moves at $0.80c$, so the trip takes $8.0/0.80 = 10$ years, and Earth clocks age 10 years. The astronaut's clock, running slow, advances only $10/\gamma = 6.0$ years.

From the astronaut's frame: the ship is at rest and it is the *distance* that is contracted:

$$L = \frac{L_p}{\gamma} = \frac{8.0\,\text{ly}}{1.67} = 4.8\,\text{ly}, \tag{35.2.7}$$

covered at $0.80c$ in $4.8/0.80 = \boxed{6.0\,\text{years}}$.

Both frames agree on what the astronaut's clock reads: they simply attribute it to different causes. Earth says the clock ran slow; the astronaut says the journey was short. This even-handedness is the heart of relativity. (A technician watching through a telescope sees the arrival later still: the arrival happens at Earth-time 10 years but $8\ \text{ly}$ away, so its light returns at year 18.)

Checkpoint 35.2

Packing for a trip at $0.99c$, a traveler wonders whether to buy smaller clothes (since they will be "thinner" from contraction) and reserve a shorter bunk (since they will be "shorter" lying down). What should they actually do, and why? *(Think about which frame the traveler is in relative to their own body.)*

Answer

Which Inertial Frame Measures the Proper Quantity?

Every time-dilation and length-contraction problem turns on one decision: *which inertial frame measures the proper quantity?* Get it right and the formulas take care of themselves; get it wrong and the answer comes out off by a factor of γ^2. The word *proper* does not mean "correct," and it does not mean "the frame that is really at rest." It comes from the French *propre*, "own": a clock's own time, an object's own length. Each definition comes with a test you can apply.

- **Proper time Δt_p.** Identify and name the two events. Proper time is measured in the frame where *both events happen at the same place*, so that one clock is present at both. Every other frame measures a longer time, $\Delta t = \gamma\ \Delta t_p$.
- **Proper length L_p.** Identify and name the object, or the two endpoints, whose length you want. Proper length is measured in the frame where *that object is at rest*. Every other frame measures a shorter length, $L = L_p/\gamma$.

A reliable routine for setting up problems:

- Write the two events in words, for example "Event 1: the ship leaves Earth. Event 2: the ship reaches the star."
- **For proper time:** find the observer for whom both events occur at the same location. Their frame measures Δt_p.
- **For proper length:** ask what is at rest in which frame. A ship is at rest in the ship's frame; the distance from Earth to a star is at rest in Earth's frame.

The most common trap is assuming that one frame measures both proper quantities. It usually does not. In the trip to Sirius, the proper *length* $(8.0\ \text{ly})$ belongs to Earth's frame, because Earth and Sirius are at rest there, while the proper *time* $(6.0$ years) belongs to the ship, because the ship is present at both departure and arrival. Table 35.1 sorts out some common situations.

Table 35.1. Which frame measures the proper time and which the proper length.

Situation	Proper time Δt_p	Proper length L_p
A muon is created high in the atmosphere and reaches the ground	the muon's frame (the muon is present at its creation and at its arrival)	Earth's frame (the thickness of the atmosphere)
A ship travels from Earth to a star	the ship's clock	Earth's frame (the Earth–star distance)
A ship passes a space station; how long does it take to go by?	the station's clock (the nose and the tail pass the *same* point on the station)	the ship's frame (the length of the ship)
A pendulum swings on a passing train	the train's frame (each swing begins and ends at the same place on the train)	—

Example 35.4: A ship passing a station

A spaceship with a proper length of $100\ \text{m}$ passes a space station at $0.80c$. (a) How long is the ship according to the station? (b) According to the station's clock, how long does the ship take to pass the station, from nose to tail? (c) How long does this take according to clocks on the ship?

Solution

At $\beta = 0.80$, $\gamma = 1/\sqrt{1-0.64} = 1/0.60 = 1.67$, and $v = 0.80c = 2.40 \times 10^8\ \text{m/s}$.

(a) Length. The ship is at rest in its own frame, so the ship's frame measures the proper length, $L_p = 100\ \text{m}$. The station measures

$$L = \frac{L_p}{\gamma} = \frac{100\,\text{m}}{1.67} = \boxed{60.0\,\text{m}}. \tag{35.2.8}$$

(b) Time on the station's clock. Event 1: the nose passes the station. Event 2: the tail passes the station. Both events happen at the station, so the station's clock measures the *proper time*. In the station's frame a $60.0\ \text{m}$ ship goes by at v:

$$\Delta t_p = \frac{L}{v} = \frac{60.0\,\text{m}}{2.40 \times 10^8\,\text{m/s}} = \boxed{0.250\,\mu\text{s}}. \tag{35.2.9}$$

(c) Time on the ship's clocks. In the ship's frame the station moves past the full $100\ \text{m}$ of ship at v:

$$\Delta t = \frac{L_p}{v} = \frac{100\,\text{m}}{2.40 \times 10^8\,\text{m/s}} = \boxed{0.417\,\mu\text{s}}, \tag{35.2.10}$$

which agrees with $\gamma\ \Delta t_p = (1.67)(0.250\ \mu\text{s})$, as it must. The station has the proper *time* but not the proper *length*; the ship has the proper length but not the proper time.

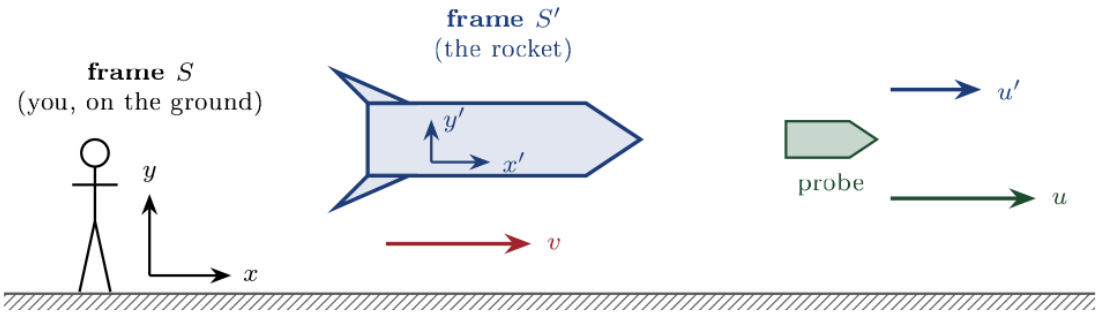
Relativity of Velocities

If a plane flies at $1000\ \text{m/s}$ and fires a missile forward at $500\ \text{m/s}$ relative to itself, common sense adds them: $1500\ \text{m/s}$. That works beautifully at low speed but cannot be exactly right: applied to light it would let a flashlight on a fast rocket beat c, violating the second postulate. The correct relativistic rule for combining velocities along a line is

Key Equation: Relativistic velocity addition

$$u' = \frac{u - v}{1 - uv/c^2}, \tag{35.4}$$

where v is the velocity of frame S' relative to S, u is the object's velocity measured in S, and u' is its velocity measured in S' (Figure 35.5). (Equivalently, $u = (u' + v)/(1 + u'v/c^2)$.) To keep the labels straight, name the three bodies (you, the rocket, the probe) and ask, for each velocity, *who is doing the measuring*. All three lie along one line, so they carry signs: take the direction of v as positive, and make u or u' negative if the object moves the other way.



v: velocity of frame S' (the rocket), measured in S
u': velocity of the probe, measured in S' (by the rocket crew)
u: velocity of the probe, measured in S (by you)

$$\text{Newton: } u = u' + v \qquad \text{Einstein: } u = \frac{u' + v}{1 + u'v/c^2} \quad \text{or} \quad u' = \frac{u - v}{1 - uv/c^2}$$

Figure 35.5. The three velocities in relativistic velocity addition. Frame S' (a rocket) moves at v relative to frame S (you, on the ground). A probe moves at u' as measured by the rocket crew, and at u as measured by you. At low speeds $u = u' + v$; the relativistic rule keeps u below c.

•Figure description

The three velocities in relativistic velocity addition. On the left an observer stands on the ground beside black x and y axes, labeled frame S, you on the ground. In the middle a blue rocket carries its own x prime and y prime axes, labeled frame S prime, the rocket; a red arrow below it is labeled v. To the right, ahead of the rocket, a small green probe has two arrows: a blue one labeled u prime and a longer green one labeled u. A key underneath reads: v, velocity of frame S prime, the rocket, measured in S; u prime, velocity of the probe, measured in S prime, by the rocket crew; u, velocity of the probe, measured in S, by you. Below that the two rules are compared: Newton, u equals u prime plus v; Einstein, u equals u prime plus v over 1 plus u prime v over c squared, or u prime equals u minus v over 1 minus u v over c squared.

The denominator is the relativistic correction; when speeds are tiny compared with c it is essentially 1 and the formula collapses to ordinary addition. But the correction guarantees that nothing built from sub-light speeds ever exceeds c, and that light combined with any speed is still exactly c.

Example 35.5: Adding large velocities

(a) A plane moves away from you at $1000\ \text{m/s}$; it launches a second plane forward at $500\ \text{m/s}$ relative to itself. How fast is the second plane moving relative to you? (b) Now let the first move at $0.80c$ and launch the second at $0.80c$ relative to itself. Find the second plane's speed relative to you.

Solution

(a) Using $u = (u' + v)/(1 + u'v/c^2)$ with $v = 1000$, $u' = 500\ \text{m/s}$, the correction $u'v/c^2 = 5 \times 10^5/9 \times 10^{16} \approx 6 \times 10^{-12}$ is negligible:

$$u = \frac{1000 + 500}{1 + 6 \times 10^{-12}} \approx \boxed{1500\,\text{m/s}}, \tag{35.2.11}$$

just as Newton would say.

(b) Now $v = 0.80c$, $u' = 0.80c$:

$$u = \frac{0.80c + 0.80c}{1 + (0.80)(0.80)} = \frac{1.60c}{1.64} = \boxed{0.976c}. \tag{35.2.12}$$

Adding $0.80c$ and $0.80c$ gives not $1.6c$ but $0.976c$: still less than c, as it must be. You can pile on speed forever and never reach light speed.

Answers to Checkpoints

Checkpoint 35.1. By Earth's clock. The *proper* time of the voyage is measured by the one clock present at both the departure and the return: the ship's clock, which travels with the crew. Every other clock measures a longer, dilated time. (For a round trip, the ship's clock is the one that turned around.) So the Earth clock reads more elapsed time, and a crew paid by it earns more.

Checkpoint 35.2. Nothing. In their own rest frame the traveler, their clothes, and their bunk are all at rest and have their ordinary lengths; there is no frame in which you move relative to yourself. Only an observer who sees the traveler moving at $0.99c$ measures their length contracted, and only along the direction of motion. Their normal clothes will fit fine.

This page titled 35.2: Consequences of Special Relativity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.3: Lorentz Transformations

Time dilation and length contraction are special cases of a single coordinate transformation between frames. In Newtonian physics, a frame S' moving at speed v along the x-axis relates to S (Figure 35.6) by the **Galilean transformation**: $x' = x - vt$, with $t' = t$; time is universal. Relativity replaces this with the **Lorentz transformation**, which mixes space and time together. For the differences between a pair of events,

Key Equation: Lorentz transformation (pairs of events)

$$\Delta x' = \gamma(\Delta x - v\,\Delta t), \qquad \Delta x = \gamma(\Delta x' + v\,\Delta t'), \tag{35.5}$$

$$\Delta t' = \gamma\left(\Delta t - \frac{v\,\Delta x}{c^2}\right), \qquad \Delta t = \gamma\left(\Delta t' + \frac{v\,\Delta x'}{c^2}\right). \tag{35.6}$$

The transverse coordinates are unchanged ($\Delta y' = \Delta y$, $\Delta z' = \Delta z$). Two features are worth dwelling on. First, when $v \ll c$ we have $\gamma \to 1$ and the $v\ \Delta x/c^2$ term vanishes, recovering the Galilean transformation exactly: relativity contains Newton's physics as its low-speed limit. Second, the time equation contains a position term: how much time elapses between two events depends on *where* they happen. That single term is the seed of the strangest consequence of all, to which we now turn.

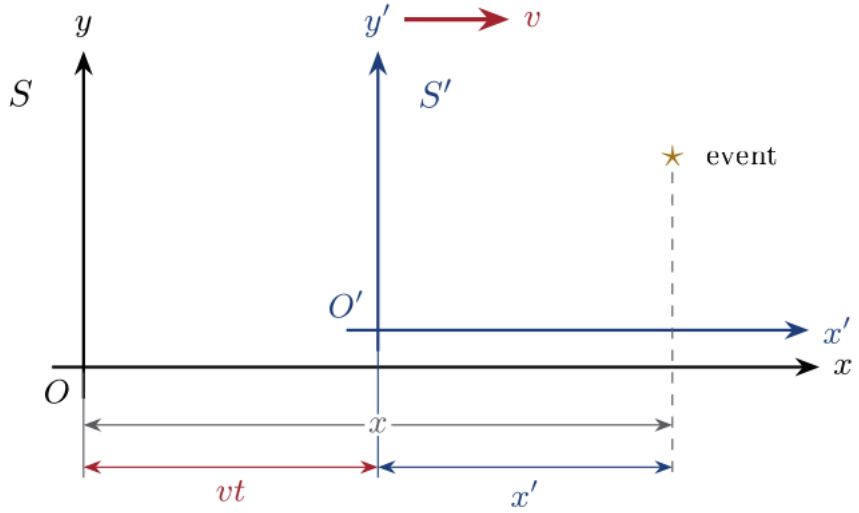


Figure 35.6. Frame S' moves at speed v along the x-axis of frame S; the origins coincide at $t = t' = 0$. One event has coordinates (x, y, t) in S and (x', y', t') in S'. Common sense (the Galilean transformation) reads $x = vt + x'$ straight off the picture, with $t' = t$; the Lorentz transformation replaces both statements. (The S' axes are drawn slightly raised only so the two x-axes can be told apart; in fact $y' = y$.)

•Figure description

Two coordinate frames. Frame S has black x and y axes with origin O. Frame S prime has blue x prime and y prime axes with origin O prime, and a red arrow shows it moving to the right at speed v relative to S. A star marks a single event up and to the right. Dashed lines drop from the event to both x axes. Dimension lines below show that the event's coordinate x in S equals the distance v t between the two origins plus the event's coordinate x prime in S prime -- the Galilean picture that the Lorentz transformation corrects.

Simultaneity

Look again at the time equation, $\Delta t' = \gamma(\Delta t - v\ \Delta x/c^2)$. Suppose two events are simultaneous in frame S, so $\Delta t = 0$, but happen at different places, $\Delta x \neq 0$. Then in S',

$$\Delta t' = -\gamma\,\frac{v\,\Delta x}{c^2} \neq 0. \tag{35.7}$$

The events are *not* simultaneous in S'. Simultaneity is not absolute; it depends on the observer's motion.

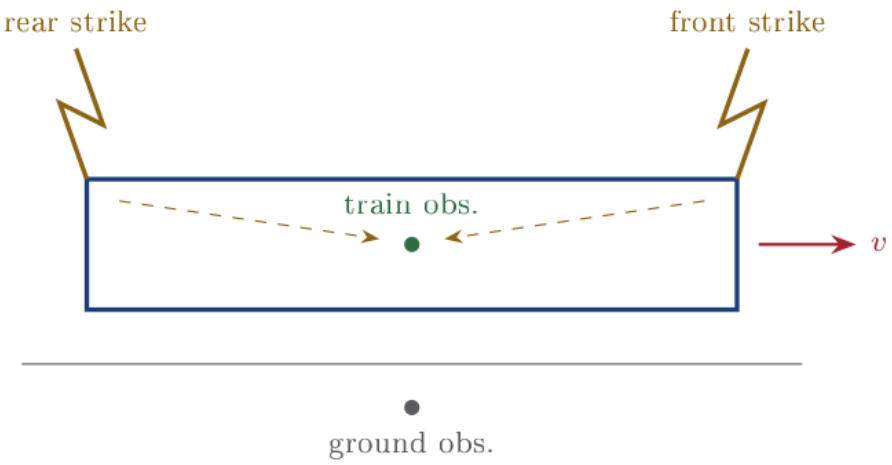


Figure 35.7. Relativity of simultaneity. Lightning strikes both ends of the moving car. The ground observer (equidistant) judges the strikes simultaneous; the train observer, moving toward the front pulse, receives it first and judges the front strike earlier. Both are correct in their own frame.

•Figure description

Einstein's train. A rectangular train car moves to the right at speed v, with an observer marked at its exact center. Gold lightning bolts strike both ends of the car at once. Dashed gold arrows show the light from each strike traveling inward toward the center observer. A second observer stands on the ground below, level with the center of the car. The ground observer is equidistant from both strike points and receives the two flashes together; the train observer is moving toward the front flash and away from the rear one, so the front flash arrives first and that strike is judged to have happened earlier.

Einstein made this vivid with a train (Figure 35.7). Lightning strikes both ends of a fast-moving train car at the instant an observer at the center of the car is level with an observer standing on the ground. The ground observer, equidistant from the two strike points, sees the flashes arrive together and rightly judges them simultaneous. But the train observer is rushing *toward* the light from the front strike and *away* from the light from the rear; since both pulses travel at c, the front flash reaches them first. The train observer concludes, just as rightly, that the front strike happened *before* the rear one. Neither is mistaken. There is no universal “now” on which all observers must agree, only each frame’s own slicing of spacetime into “now” and “then.” This is the deepest break with Newton: he assumed a single cosmic clock ticking for everyone, and there simply isn’t one.

Checkpoint 35.3

(a) Two events happen at the *same place* and the same time in one frame. Can an observer in another inertial frame judge them to be non-simultaneous? Use equation (35.6). (b) In Figure 35.7, which observer is right about whether the strikes were simultaneous?

Answer

Answers to Checkpoints

Checkpoint 35.3. (a) No. With $\Delta x = 0$ and $\Delta t = 0$ in one frame, equation (35.6) gives $\Delta t' = \gamma(0 - 0) = 0$ in every other frame. The relativity of simultaneity applies only to events that are separated in space, since the offending term is $v\ \Delta x/c^2$. Events at the same place and time are simultaneous for everyone. (b) Both. Each observer’s judgment is correct in that observer’s own frame. Simultaneity for separated events is not absolute; it depends on the frame, and no frame is preferred.

This page titled 35.3: Lorentz Transformations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.4: Doppler Effect for Light

Like sound, light shifts in frequency when source and observer move relative to one another, but the relativistic formula differs from the sound version because there is no medium and because time dilation contributes. Working in wavelength, which is how light is usually measured, the observed wavelength relates to the emitted (rest-frame) wavelength by

Key Equation: Relativistic Doppler effect

$$\lambda_{\text{obs}} = \lambda_{\text{src}}\sqrt{\frac{1 \pm \beta}{1 \mp \beta}}, \qquad \beta = \frac{v}{c}. \tag{35.8}$$

The sign choice follows the motion: for a *receding* source the wavelength stretches (the upper sign, $\sqrt{(1+\beta)/(1-\beta)}$), a **redshift**; for an *approaching* source it compresses: a **blueshift**. This is one of astronomy's most powerful tools. Edwin Hubble found that the light from distant galaxies is uniformly redshifted, and the farther the galaxy, the greater the shift: the discovery that the universe is expanding.

Example 35.6: A receding galaxy

A hydrogen line emitted at $650\ \text{nm}$ is observed at $1458\ \text{nm}$ in light from a distant galaxy. How fast is the galaxy receding?

Solution

The wavelength is stretched, so the galaxy recedes; use the upper sign. Let $R = \lambda_{\text{obs}}/\lambda_{\text{src}} = 1458/650 = 2.243$. Then

$$R^2 = \frac{1+\beta}{1-\beta} = 5.03 \;\Rightarrow\; 1+\beta = 5.03(1-\beta) \;\Rightarrow\; \beta = \frac{R^2-1}{R^2+1} = \frac{5.03-1}{5.03+1} = 0.668. \tag{35.4.1}$$

The galaxy recedes at

$$v = 0.668c \approx \boxed{2.0 \times 10^8\ \text{m/s}}. \tag{35.4.2}$$

Such large redshifts are routine for distant galaxies, and quantifying them is how we map the expansion, and the history, of the cosmos.

This page titled 35.4: Doppler Effect for Light is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.5: Momentum and Energy

If velocities do not add the old way, momentum cannot be the old mv either, or it would not be conserved in every frame. The relativistic **momentum** of a particle of mass m is

$$\mathbf{p} = \gamma m \mathbf{v}. \tag{35.9}$$

As $v \to c$, $\gamma \to \infty$ and so does p: an infinite push would be needed to reach light speed, which is why no massive object ever does. Tracking the work needed to accelerate a particle leads to relativity's most famous result: that mass itself is a form of energy. A particle at rest has **rest energy**

$$E_R = mc^2, \tag{35.10}$$

and a moving particle has **total energy** and **kinetic energy**

Key Equation: Relativistic energy

$$E = \gamma mc^2 = E_R + K, \qquad K = (\gamma - 1)mc^2, \tag{35.11}$$
$$E^2 = (pc)^2 + (mc^2)^2. \tag{35.12}$$

The kinetic energy is the *excess* over rest energy; expand $(\gamma - 1)mc^2$ for small v and it reduces to the familiar $\frac{1}{2}mv^2$. The last relation, $E^2 = (pc)^2 + (mc^2)^2$, ties energy and momentum together and even covers massless particles: setting $m = 0$ gives $E = pc$, exactly the energy–momentum relation a photon obeys. Because the numbers in SI units are tiny and awkward, particle physicists measure energy in electron-volts and mass in MeV/c^2, using $c^2 = 931.494\ \mathrm{MeV/u}$.

Example 35.7: The energetics of a proton

A proton has mass m_p. (a) Find its rest energy in electron-volts. (b) If its total energy is three times its rest energy, find its speed. (c) Find its kinetic energy. (d) Find its momentum.

Solution

(a) The proton's rest energy is a standard value,

$$E_R = m_p c^2 = \boxed{938\,\mathrm{MeV}}\ (= 1.50 \times 10^{-10}\,\mathrm{J}). \tag{35.5.1}$$

(b) If $E = 3E_R$, then $\gamma = E/E_R = 3$. Solving $\gamma = 1/\sqrt{1-\beta^2} = 3$:

$$1 - \beta^2 = \tfrac{1}{9} \ \Rightarrow\ \beta = \sqrt{\tfrac{8}{9}} = 0.943, \qquad v = \boxed{0.943c}. \tag{35.5.2}$$

(c) The kinetic energy is the total minus the rest energy:

$$K = E - E_R = 3E_R - E_R = 2E_R = 2(938\,\mathrm{MeV}) = \boxed{1876\,\mathrm{MeV}}. \tag{35.5.3}$$

(d) From $E^2 = (pc)^2 + E_R^2$ with $E = 3E_R$,

$$(pc)^2 = E^2 - E_R^2 = 9E_R^2 - E_R^2 = 8E_R^2 \ \Rightarrow\ pc = \sqrt{8}\,E_R = 2653\,\mathrm{MeV}, \tag{35.5.4}$$

so $p = \boxed{2653\,\mathrm{MeV}/c}$.

Checkpoint 35.4

(a) A particle's speed doubles, from $0.4c$ to $0.8c$. Does its kinetic energy double, more than double, or less? (b) Use $K = (\gamma - 1)mc^2$ to explain why no amount of work can bring a massive particle up to the speed of light. (c) A photon has no mass. According to equation (35.12), how are its energy and momentum related?

Answer

Answers to Checkpoints

Checkpoint 35.4. (a) More than double. At $0.4c$, $\gamma = 1/\sqrt{1-0.16} = 1.091$, so $K = 0.091\ mc^2$; at $0.8c$, $\gamma = 1/\sqrt{1-0.64} = 1.667$, so $K = 0.667\ mc^2$: more than seven times larger. (Classically the factor would be exactly 4.) (b) As $v \to c$, $\gamma = 1/\sqrt{1-v^2/c^2}$ grows without bound, and so does $K = (\gamma - 1)mc^2$. Reaching c would take infinite kinetic energy, and so infinite work. (c) With $m = 0$, $E^2 = (pc)^2$, so $E = pc$: a photon's energy and momentum are proportional, with c as the constant.

This page titled 35.5: Momentum and Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.6: General Relativity

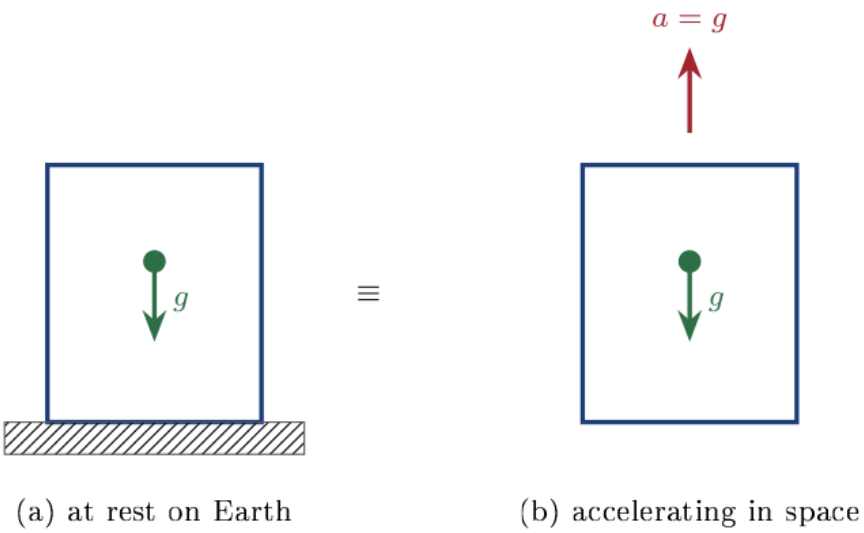


Figure 35.8. The equivalence principle. A person in a closed box cannot tell, by any local experiment, whether (a) the box rests in a gravitational field g or (b) accelerates through empty space at $a = g$. Gravity and acceleration are locally indistinguishable.

•Figure description

The equivalence principle, shown as two boxes with an equals-equivalent sign between them. In panel (a) a closed elevator rests on hatched ground representing the Earth, and a ball released inside falls toward the floor, its acceleration shown by an arrow labeled g. In panel (b) an identical elevator floats in empty space with a red arrow above it showing it is being accelerated upward at a equals g, and a ball released inside falls toward the floor with an identical arrow labeled g. Nothing inside either box could tell the two situations apart.

Special relativity is restricted to inertial frames: the constant-velocity case we agreed to at the start. In 1915 Einstein removed that restriction and, in doing so, produced a theory of gravity. Its cornerstone is the **equivalence principle**: *no local experiment can distinguish a uniformly accelerating frame from a frame at rest in a gravitational field.* Stand in a sealed elevator and feel a force pressing you to the floor; you cannot tell, from inside, whether the elevator sits on Earth or is being accelerated upward at g in deep space (Figure 35.8). Gravity and acceleration are locally the same thing.

From this seed grow general relativity's predictions, several now part of daily life:

- **Gravitational time dilation:** clocks run slower deeper in a gravitational well. Clocks at sea level tick slightly slower than clocks on a mountaintop. The GPS satellites must correct for this (and for their orbital speed) or navigation would drift by kilometers a day.
- **Bending of light:** mass curves the paths of light rays. Starlight grazing the Sun is deflected (famously confirmed during the 1919 eclipse) and galaxies act as *gravitational lenses*, warping the images of objects behind them.
- **Black holes and gravitational waves:** where mass is concentrated enough, spacetime curves so steeply that not even light escapes; and accelerating masses radiate ripples in spacetime, the gravitational waves first detected directly in 2015 (Chapter 33).

In Einstein's picture, mass and energy tell spacetime how to curve, and curved spacetime tells matter how to move. What we feel as gravity is matter following the straightest available path through a geometry bent by mass. That is a long way from a car waiting at a curb, but it began with the same stubborn question Einstein posed to Newton: prove who is moving. The answer, in the end, was that motion, time, space, and even gravity are relationships between observers and events, not a fixed stage on which physics plays out.

This page titled 35.6: General Relativity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.S: Relativity (Summary)

Summary

- **Postulates:** the laws of physics are identical in all inertial frames, and light travels at c in every inertial frame. The Michelson–Morley null result motivated the second.
- **Lorentz factor:** $\gamma = 1/\sqrt{1 - v^2/c^2} \geq 1$; near 1 at everyday speeds, diverging as $v \to c$.
- **Time dilation:** $\Delta t = \gamma\ \Delta t_p$, moving clocks run slow; Δt_p is the proper time. **Length contraction:** $L = L_p/\gamma$, moving objects shorten along the motion.
- **Proper frames:** the proper time is measured where both events happen at the same place (one clock present at both); the proper length is measured where the object is at rest. They are often different frames.
- **Velocity addition:** $u' = (u - v)/(1 - uv/c^2)$, which keeps all speeds $\leq c$. **Lorentz transformation:** $\Delta x' = \gamma(\Delta x - v\Delta t)$, $\Delta t' = \gamma(\Delta t - v\Delta x/c^2)$, reducing to Galilean at low v.
- **Simultaneity is relative:** events simultaneous in one frame ($\Delta t = 0$, $\Delta x \neq 0$) are not in another.
- **Doppler effect:** $\lambda_{\text{obs}} = \lambda_{\text{src}}\sqrt{(1 \pm \beta)/(1 \mp \beta)}$; receding sources redshift, the signature of cosmic expansion.
- **Momentum and energy:** $p = \gamma mv$; $E = \gamma mc^2 = E_R + K$ with rest energy $E_R = mc^2$ and $K = (\gamma - 1)mc^2$; and $E^2 = (pc)^2 + (mc^2)^2$.
- **General relativity:** the **equivalence principle** makes gravity and acceleration locally indistinguishable, predicting gravitational time dilation, light bending, black holes, and gravitational waves.

Key Equations

Lorentz factor

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1 - \beta^2}}, \qquad \beta = \frac{v}{c} \tag{35.S.1}$$

Time dilation

$$\Delta t = \gamma\, \Delta t_p \qquad (\Delta t_p = \text{proper time}) \tag{35.S.2}$$

Length contraction

$$L = \frac{L_p}{\gamma} = L_p\sqrt{1 - v^2/c^2} \qquad (\text{along the motion only}) \tag{35.S.3}$$

Relativistic velocity addition

$$u' = \frac{u - v}{1 - uv/c^2}, \qquad u = \frac{u' + v}{1 + u'v/c^2} \tag{35.S.4}$$

Lorentz transformation

$$\begin{aligned} \Delta x' &= \gamma(\Delta x - v\,\Delta t) \\ \Delta t' &= \gamma\left(\Delta t - \frac{v\,\Delta x}{c^2}\right) \end{aligned} \tag{35.S.5}$$

Relativity of simultaneity

$$\Delta t = 0,\ \Delta x \neq 0 \implies \Delta t' = -\gamma\,\frac{v\,\Delta x}{c^2} \neq 0 \tag{35.S.6}$$

Relativistic Doppler effect

$$\lambda_{\text{obs}} = \lambda_{\text{src}} \sqrt{\frac{1 \pm \beta}{1 \mp \beta}} \qquad \text{(upper sign: receding, redshift)} \tag{35.S.7}$$

Relativistic momentum

$$\mathbf{p} = \gamma m \mathbf{v} \tag{35.S.8}$$

Rest, total and kinetic energy

$$\begin{aligned} E_R &= mc^2 \\ E &= \gamma mc^2 = E_R + K \\ K &= (\gamma - 1)mc^2 \end{aligned} \tag{35.S.9}$$

Energy–momentum relation

$$E^2 = (pc)^2 + (mc^2)^2 \qquad (m = 0 \implies E = pc) \tag{35.S.10}$$

This page titled 35.S: Relativity (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

35.E: Relativity (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Take $c = 3.00 \times 10^8\ \text{m/s}$ and the proton rest energy as $938\ \text{MeV}$. Recall $\gamma = 1/\sqrt{1 - v^2/c^2}$.

Sections 35.1–35.2 — Foundations

35.1 ★ *[Conceptual.]* State Einstein's two postulates of special relativity, and explain how the Michelson–Morley experiment motivated the second.

35.2 ★★ Compute the Lorentz factor γ for an object moving at (a) $0.10c$, (b) $0.50c$, (c) $0.90c$, and (d) $0.99c$. Comment on how rapidly γ grows.

Section 35.2.1 — Time Dilation

35.3 ★ A spaceship clock measures a process taking $4.0\ \text{s}$ as the ship passes at $0.80c$. How long does the process take in the Earth frame?

35.4 ★★ A muon at rest decays in $2.2\ \mu\text{s}$ on average. If it travels at $0.995c$, what is its average lifetime in the lab frame, and how far does it travel in that time?

35.5 ★★★ *[Real-world application: navigation.]* An atomic clock is flown around the world at an average $250\ \text{m/s}$ for 40 hours (ignore gravitational effects). By how much does it lag a ground clock due to special-relativistic time dilation?

35.6 ★★ Charged pions have a proper mean lifetime of $26\ \text{ns}$. In a pion beam at an accelerator, the pions travel an average of $15\ \text{m}$ in the lab before they decay.

a. Find the pions' speed as a fraction of c. (Hint: the distance is $d = v\gamma\tau$, so $d/(c\tau) = \beta\gamma = \beta/\sqrt{1-\beta^2}$; solve for β.)
b. Find γ.
c. Find the mean lifetime of the pions as measured in the lab.

Sections 35.2.2–35.2.3 — Length Contraction and Proper Frames

35.7 ★ A spaceship has a proper length of $120\ \text{m}$. How long does it appear to an observer past whom it travels at $0.60c$?

35.8 ★★ At what speed must an object move for its length along the motion to be contracted to 50% of its proper length?

35.9 ★★★ A starship travels to a star 4.2 light-years away (Earth frame) at $0.90c$. (a) How long does the trip take in the Earth frame? (b) What distance and trip time does the crew measure?

35.10 ★ *[Conceptual.]* A high-speed train passes a station without stopping. For each measurement below, identify and name the two events or the object involved, and state which inertial frame (the train's or the ground's) measures the proper time or proper length: (a) the time between two ticks of a passenger's wristwatch; (b) the length of one train car; (c) the time for the entire train to pass a signal post beside the track; (d) the distance between two stations on the line.

35.11 ★★ An astronaut on a ship moving at $0.80c$ relative to Earth counts her pulse at 72 beats per minute on the ship's clock. (a) Which inertial frame measures the proper time between her heartbeats? Explain. (b) What is the time between heartbeats as measured from Earth, and how many beats per minute is that? (c) The ship is $90\ \text{m}$ long in its own frame. How long is it according to Earth? (d) A beacon on Earth flashes once every $1.00\ \text{s}$ by Earth clocks. Which frame measures the proper time between flashes, and what interval does the astronaut measure?

35.12 ★★ *[Real-world application: cosmic rays.]* A muon is created by a cosmic ray $12.0\ \text{km}$ above the ground (Earth frame) and travels straight down at $0.995c$. (a) How long does the trip to the ground take in Earth's frame? (b) Identify and name the two events, and decide which inertial frame measures the proper time between them. How long does the trip take by the muon's own

clock? (c) Which frame measures the proper length of the 12.0-km stretch of atmosphere? How thick is that stretch in the muon's frame? (d) Show that your answers to (b) and (c) are consistent with the muon seeing the ground approach at $0.995c$.

Sections 35.2.4–35.3 — Velocities & Lorentz Transformations

35.13 ★★ A rocket moves at $0.60c$ relative to Earth and fires a probe forward at $0.50c$ relative to the rocket. Find the probe's speed relative to Earth.

35.14 ★★ Two spaceships approach Earth from opposite directions, each at $0.70c$ as measured from Earth. How fast does one ship measure the other to be approaching?

35.15 ★★★ In frame S, two events occur $600\ \text{m}$ apart and $2.0\ \mu\text{s}$ apart. Frame S' moves at $0.80c$ along the line joining them. Find the spatial and time separations of the events in S'.

Section 35.3.1 — Simultaneity

35.16 ★★ *[Conceptual.]* Two events are simultaneous in your frame but occur at different locations. Explain why a moving observer will, in general, not find them simultaneous, and what determines which event they judge to happen first.

Section 35.4 — Doppler Effect for Light

35.17 ★★ A spectral line emitted at $486\ \text{nm}$ is observed at $456\ \text{nm}$ from an approaching star. Find the star's speed.

35.18 ★★★ *[Real-world application: traffic.]* A driver claims a red traffic light ($\lambda = 650\ \text{nm}$) appeared green ($\lambda = 520\ \text{nm}$) because the car was approaching it. (a) What speed would be required? (b) Comment on the plausibility of the excuse.

Section 35.5 — Momentum and Energy

35.19 ★ Find the rest energy of an electron ($m = 9.11 \times 10^{-31}\ \text{kg}$) in MeV.

35.20 ★★ An electron is accelerated to $0.99c$. Find its total energy and kinetic energy in MeV (electron rest energy $0.511\ \text{MeV}$).

35.21 ★★★ A particle's total energy is 5.0 times its rest energy. Find (a) its speed and (b) the ratio of its momentum to mc.

Section 35.6 — General Relativity & Synthesis

35.22 ★★ *[Conceptual.]* State the equivalence principle and give one everyday technology whose accuracy depends on a general-relativistic correction.

35.23 ★★★ *[Conceptual — "prove it."]* Return to the chapter opener, in which Einstein challenges Newton to prove who is really moving. Using the first postulate, explain why the challenge cannot be met for two observers in uniform relative motion, and identify what *does* break the symmetry in the twin/Sirius scenario, where the traveler genuinely ages less.

Challenge Problems

35.24 ★★★ *[Real-world application: navigation — why GPS needs relativity.]* A GPS satellite orbits at an altitude of $2.02 \times 10^7\ \text{m}$ above Earth's surface ($R_E = 6.371 \times 10^6\ \text{m}$) with an orbital speed of $3874\ \text{m/s}$. Two relativistic effects act on its atomic clock in *opposite* directions: special-relativistic time dilation from its speed, and the general-relativistic speeding up of a clock higher in a gravitational field. For weak fields and slow speeds the two fractional rate changes add:

$$\frac{\Delta f}{f} = \underbrace{-\frac{v^2}{2c^2}}_{\text{speed}} + \underbrace{\frac{GM}{c^2}\left(\frac{1}{R_E} - \frac{1}{r}\right)}_{\text{altitude}}, \tag{35.E.1}$$

with $GM = 3.986 \times 10^{14}\ \mathrm{m^3/s^2}$ and $r = R_E + h$.

a. Evaluate the speed term. By how many microseconds per day does this alone make the satellite clock run *slow* relative to a ground clock?
b. Evaluate the altitude term. By how many microseconds per day does this alone make it run *fast*?
c. Combine them. What is the net drift, in microseconds per day, and in which direction?
d. A GPS receiver finds its position by timing signals that travel at c. If the net drift in (c) went uncorrected, how large a position error would accumulate in one day? Comment on whether a navigation system could function without accounting for it.

35.25 ★★★ *[Real-world application: particle physics — protons in the LHC.]* The Large Hadron Collider accelerates protons to a total energy of $6.5\ \mathrm{TeV}$ (the energy of its 2015–2018 run) around a ring of circumference $26.659\ \mathrm{km}$. The proton rest energy is $mc^2 = 938.272\ \mathrm{MeV}$.

a. Find the Lorentz factor γ of these protons.
b. For $\gamma \gg 1$, show that $1 - \beta \approx 1/(2\gamma^2)$. Evaluate it, and find by how many meters per second the protons fall short of c.
c. In the rest frame of a proton, the whole $26.659\ \mathrm{km}$ ring is length-contracted. How long is it? Compare with something of everyday size.
d. Find the time for one revolution in the laboratory frame, and the proper time experienced by the proton for that same revolution. Explain briefly how these two answers are consistent with the ring's contracted length in (c).

35.26 ★★★ *[Real-world application: astrophysics — apparent superluminal motion.]* Astronomers tracking blobs of plasma ejected from the cores of active galaxies have repeatedly measured apparent transverse speeds *greater* than c. Nothing is actually exceeding c; the illusion arises because the blob is chasing its own light. A blob moves at speed $v = \beta c$ along a line making angle θ with the line of sight toward Earth.

a. In a time Δt the blob moves a transverse distance $v\Delta t \sin\theta$ and also moves $v\Delta t \cos\theta$ *closer* to us, so the second photon has less distance to travel. Show that the apparent elapsed time on Earth is $\Delta t_{\mathrm{obs}} = \Delta t\ (1 - \beta\cos\theta)$, and hence that the apparent transverse speed is

$$v_{\mathrm{app}} = c\,\frac{\beta \sin\theta}{1 - \beta\cos\theta}. \tag{35.E.2}$$

b. Evaluate v_{app}/c for $\beta = 0.98$ and $\theta = 15^\circ$.
c. Show that for a given β the apparent speed is largest when $\cos\theta = \beta$, and that its maximum value is $\gamma\beta\ c$. Evaluate both the angle and the maximum for $\beta = 0.98$.
d. Explain in two or three sentences why no causality problem arises here, and what an observation of $v_{\mathrm{app}} = 5c$ lets you conclude about the true speed of the blob.

This page titled 35.E: Relativity (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36: TEMPERATURE AND KINETIC THEORY OF GASES

Open a bottle of perfume on one side of a quiet room, and someone on the far side may not smell it for a minute or more. That seems unremarkable until you learn how fast the molecules of the air are moving. At room temperature a typical nitrogen molecule travels at about $470\ \mathrm{m/s}$: faster than the speed of sound, and more than half the cruising speed of a jetliner. A perfume molecule moving at that speed should cross a five-meter room in about a hundredth of a second. Why, then, does the scent take so long to arrive?

The answer is that no molecule in the air travels in a straight line for long. Each one collides with a neighbor billions of times every second, so its path is a tangled random walk with steps only a few tens of nanometers long. By the end of this chapter you will be able to calculate both numbers yourself, the molecular speed and the distance between collisions, from nothing more than the temperature and pressure of the air.

This chapter is the first of three on **thermal physics**. Here we ask what temperature is and what it does to matter: how it is measured, why things expand when they warm, and how the pressure, volume, and temperature of a gas are tied together. The chapter ends with the **kinetic theory of gases**, which explains all of this in terms of molecules in motion and shows that temperature is, at bottom, a measure of molecular kinetic energy. Chapter 37 then follows energy as it flows as heat and work, and Chapter 38 asks which way that flow is allowed to go.

Learning Objectives

- State the zeroth law of thermodynamics and explain how a thermometer works.
- Convert among the Celsius, Fahrenheit, and Kelvin scales, and explain why absolute zero is the natural zero of temperature.
- Calculate the linear, area, and volume thermal expansion of solids and liquids, and the thermal stress in a constrained rod.
- Use the ideal gas law, $PV = nRT = Nk_BT$, to relate the pressure, volume, temperature, and amount of a gas.
- Derive the pressure of a gas from molecular collisions, and connect temperature to the average molecular kinetic energy, $\overline{K} = \frac{3}{2}k_BT$.
- Use the equipartition theorem, the Maxwell speed distribution, and the mean free path to describe the molecules of a gas.

We will return to the slow perfume once we have the mean free path, and again in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

36.1: Temperature
36.2: Thermal Expansion
36.3: Ideal Gas Law
36.4: Kinetic Theory of Gases
36.S: Temperature and Kinetic Theory of Gases (Summary)
36.E: Temperature and Kinetic Theory of Gases (Problems)

This page titled 36: Temperature and Kinetic Theory of Gases is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.1: Temperature

We all have a sense of hot and cold, but touch is an unreliable judge of temperature. On a cold morning a metal door handle feels much colder than the wooden door around it, although the two have been sitting side by side all night and are at the same temperature. Your skin senses how fast energy leaves your hand, not the temperature of what you touch. Physics needs something better.

Start with what happens when two objects are put in contact. A hot cup of coffee left on a table cools, and the air around it warms slightly, until the two stop changing. Objects that can exchange energy this way are in **thermal contact**; when the exchange has stopped and nothing measurable is changing any more, they are in **thermal equilibrium**. Temperature is the property that two objects in thermal equilibrium share: objects at the same temperature exchange no net energy when they touch, and objects at different temperatures do.

That idea is so basic that it is enshrined as a law of its own, the **zeroth law of thermodynamics**:

Key Equation: Zeroth law of thermodynamics

If objects A and B are each in thermal equilibrium with a third object C, then A and B are in thermal equilibrium with each other.

The name is a historical accident. The first and second laws were already numbered when physicists realized that this even more basic statement had been taken for granted, so it was slipped in ahead of them. The zeroth law is what makes a **thermometer** meaningful. Object C is the thermometer: it is brought into contact with A and allowed to come to equilibrium, and then with B. If it reads the same in both, A and B are at the same temperature, even though they have never touched.

Any property of matter that changes steadily with temperature can be made into a thermometer: the length of a column of liquid in a glass tube, the electrical resistance of a platinum wire (the platinum resistance thermometer of Chapter 20), the voltage across the junction of two different metals, or the color of the light a hot object emits. Each needs to be calibrated against fixed physical points, such as the freezing and boiling points of water.

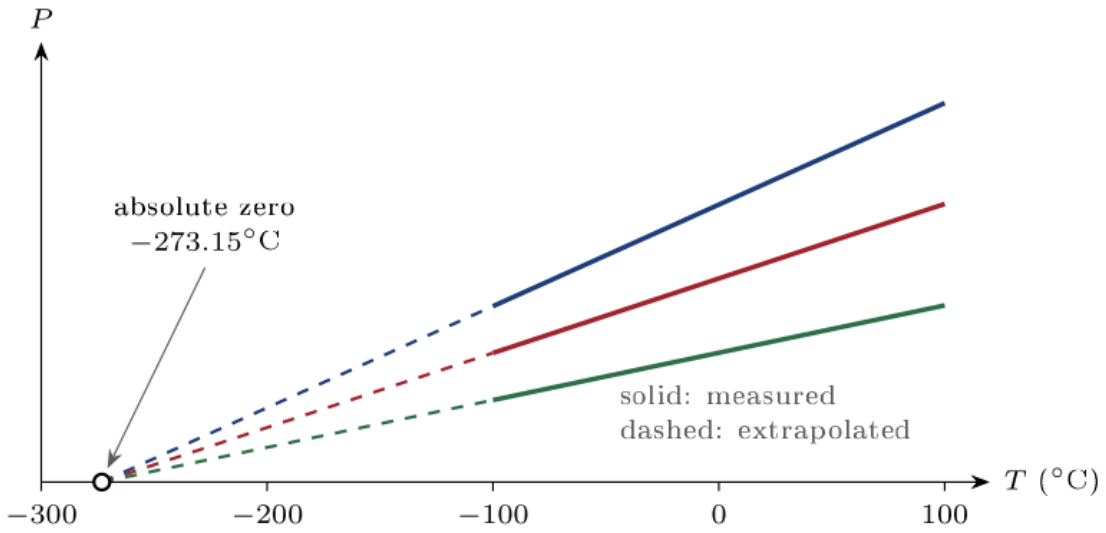


Figure 36.1. A constant-volume gas thermometer. The pressure of a fixed volume of gas rises in a straight line with temperature. Different gases, or different amounts of the same gas, give lines of different slope, but every line extrapolates to zero pressure at the same temperature, $-273.15^{\circ}\mathrm{C}$. That common point is absolute zero.

•Figure description

A graph of pressure on the vertical axis against temperature in degrees Celsius on the horizontal axis, which runs from minus 300 to plus 100. Three solid straight lines show the pressure of three different gas samples held at constant volume, measured between about minus 100 and plus 100 degrees Celsius. The lines have different slopes. Each is extended to lower temperatures as a dashed line, and all three dashed lines meet the temperature axis, where the pressure is zero, at the same point, marked minus 273.15 degrees Celsius, absolute zero.

One kind of thermometer turns out to be special. Seal a fixed volume of a dilute gas in a bulb and record its pressure at several temperatures. The points fall on a straight line (Figure 36.1). Repeat with a different gas, or a different amount of gas, and you get a different straight line. But extend every one of these lines to lower temperatures, and all of them reach zero pressure at the same point: $-273.15^{\circ}\mathrm{C}$. A negative pressure is meaningless, so no gas can be cooled below this temperature. It is the **absolute zero** of temperature, and its existence hints at what temperature really measures. As we will show in Section 36.4, the temperature of a gas is proportional to the average kinetic energy of its molecules, and absolute zero is where that kinetic energy would reach its minimum. (Quantum mechanics, Chapter 40, adds that the motion never stops completely, even there.) The coldest air temperature measured at Earth's surface is about $-89^{\circ}\mathrm{C}$, in Antarctica, but laboratories have cooled atoms to within a billionth of a kelvin of absolute zero.

Checkpoint 36.1

A thermometer placed in a bucket of water reads $25^{\circ}\mathrm{C}$. The same thermometer, moved to a block of copper, also reads $25^{\circ}\mathrm{C}$. If the copper block is dropped into the bucket, will there be any net flow of energy between them? Which law tells you so?

Answer

Temperature Scales

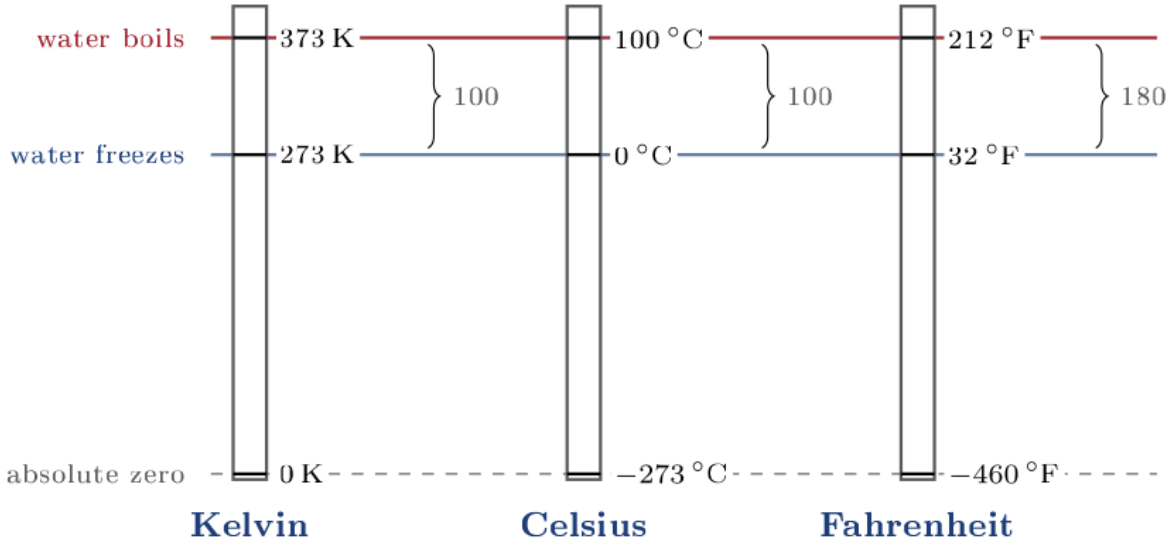


Figure 36.2. The Kelvin, Celsius, and Fahrenheit scales compared. Water freezes and boils at fixed physical points; the scales merely assign different numbers to them. Between freezing and boiling there are 100 kelvins and 100 Celsius degrees but 180 Fahrenheit degrees: Kelvin and Celsius degrees are equal in size, so a temperature *difference* is the same number in either.

•Figure description

Three vertical thermometer columns side by side, labeled Kelvin, Celsius and Fahrenheit, with three horizontal reference lines drawn across all of them and labeled at the left. The top line, water boils, reads 373 K, 100 degrees C and 212 degrees F. The middle line, water freezes, reads 273 K, 0 degrees C and 32 degrees F. The bottom dashed line, absolute zero, reads 0 K, minus 273 degrees C and minus 460 degrees F. A brace beside each column spans the gap between freezing and boiling: it is labeled 100 on the Kelvin and Celsius columns and 180 on the Fahrenheit column.

Three temperature scales are in common use (Figure 36.2). The **Fahrenheit** scale (°F) is still everyday in the United States. The **Celsius** scale (°C) sets 0° at the freezing point of water and 100° at its boiling point (both at standard atmospheric pressure). The **Kelvin** scale (K), the SI scale and the one we use in nearly all calculations, places its zero at absolute zero, so it never goes negative. Kelvin temperatures are written without a degree sign: water freezes at $273.15\ \mathrm{K}$, read "273.15 kelvins." Celsius and Kelvin degrees are the same size; the scales differ only by an offset. The conversions are

Key Equation: Temperature conversions

$$T_C = \tfrac{5}{9}(T_F - 32), \tag{36.1}$$

$$T = T_C + 273.15. \tag{36.2}$$

Here T with no subscript is the Kelvin temperature, the one that belongs in every formula of the next three chapters. Because a change of one Celsius degree equals a change of one kelvin, any *temperature difference* ΔT has the same numerical value in °C and in K, a fact we will lean on constantly. A temperature *reading* is a different matter: $20^\circ\mathrm{C}$ is $293\ \mathrm{K}$, not $20\ \mathrm{K}$.

Note: Defining the kelvin

Since 2019 the kelvin, like the other SI base units (Chapter 1), has been defined by fixing the value of a fundamental constant. Boltzmann's constant, which we meet in Section 36.3, is set to exactly $k_B = 1.380649 \times 10^{-23}\ \mathrm{J/K}$. One kelvin is then the temperature change that raises the typical energy of a molecule by a fixed, known number of joules. Before 2019 the kelvin was tied to a property of water (its triple point, where ice, liquid water, and vapor coexist, was defined to be exactly $273.16\ \mathrm{K}$).

Example 36.1: Liquid nitrogen

Liquid nitrogen, used to freeze tissue samples and to cool superconducting magnets, boils at $77.4\ \mathrm{K}$ at atmospheric pressure. Express this temperature in degrees Celsius and in degrees Fahrenheit.

Solution

Subtract the offset to get Celsius:

$$T_C = T - 273.15 = 77.4 - 273.15 = \boxed{-195.8^\circ\text{C}}. \tag{36.1.1}$$

Then invert equation (36.1), $T_F = \frac{9}{5}T_C + 32$:

$$T_F = \tfrac{9}{5}(-195.8) + 32 = -352.4 + 32 = \boxed{-320.4^\circ\text{F}}. \tag{36.1.2}$$

Liquid nitrogen is only $77\ \text{K}$ above absolute zero, but in everyday units it is hundreds of degrees below anything in ordinary experience.

Answers to Checkpoints

Checkpoint 36.1. No. The water and the copper are each in thermal equilibrium with the thermometer, so by the zeroth law they are in thermal equilibrium with each other: they are at the same temperature, and no net energy flows when they touch.

This page titled 36.1: Temperature is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.2: Thermal Expansion

Nearly every substance expands when it is heated. Bridges are built with expansion joints, gaps that let the roadway lengthen on a hot afternoon without buckling; power lines sag lower in summer than in winter; and a metal jar lid stuck fast can be loosened by running hot water over it, because the metal lid expands more than the glass jar.

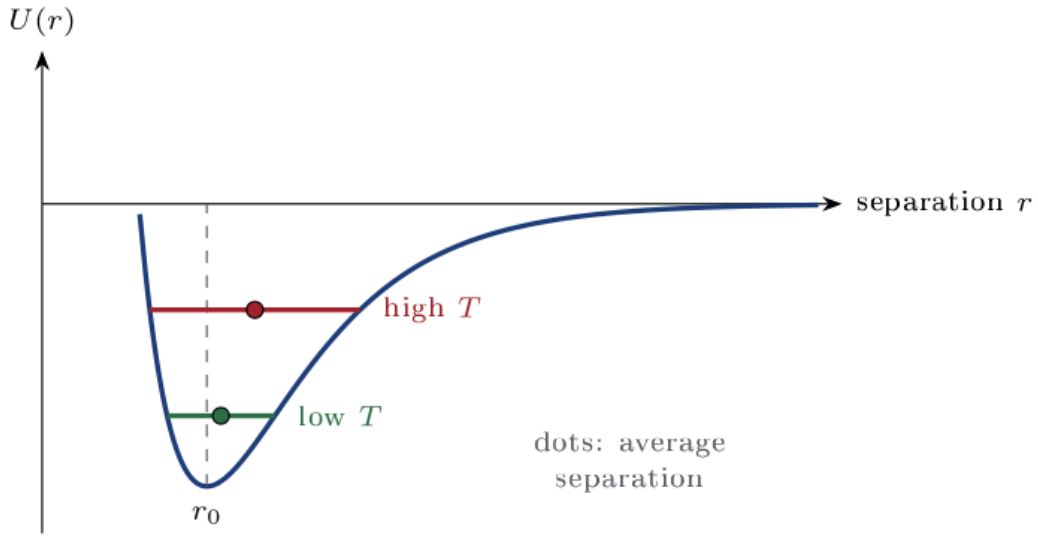


Figure 36.3. Why solids expand. The potential energy of two neighboring atoms is a lopsided well, much steeper on the near side than on the far side. As the temperature rises the atoms vibrate with more energy, and because the well is lopsided, their *average* separation (dots) moves outward. A symmetric well would produce no expansion at all.

•Figure description

A graph of the potential energy of a pair of neighboring atoms against their separation r. The curve is a lopsided well: it rises very steeply at small separations, has a minimum at the equilibrium separation r sub 0, and rises only gently and levels off at large separations. Two horizontal lines across the well mark two vibration energies, a lower one labeled low T and a higher one labeled high T. A dot marks the midpoint of each line, the average separation during the vibration. The midpoint of the low-energy line sits just to the right of r sub 0, and the midpoint of the high-energy line sits farther to the right: at higher temperature the average separation is larger.

The reason is microscopic. The atoms of a solid are held near their neighbors by forces that act like stiff springs, and they vibrate about their equilibrium positions. The potential energy of a pair of neighboring atoms is not a symmetric bowl, though: it rises very steeply if the atoms are pushed together and only gently if they are pulled apart (Figure 36.3). As the temperature rises the vibration grows, and because the well is lopsided, the atoms spend more of their time farther apart than closer together. The average spacing increases, and the whole object grows. (We look at where this potential energy curve comes from in Chapter 42.)

For a rod or any other long object, the change in length is proportional to the original length and to the temperature change:

Key Equation: Linear thermal expansion

$$\Delta L = \alpha L_0 \, \Delta T. \tag{36.3}$$

The constant α is the **coefficient of linear expansion**, a property of the material with units of K^{-1} (Table 36.1). It is small: a steel rod one meter long grows by only about $0.012 \ \mathrm{mm}$ for each kelvin of warming. Equation (36.3) applies to every linear dimension of an object, its length, width, and thickness alike, and to the size of any hole in it.

Table 36.1. Coefficients of thermal expansion near room temperature. Solids are given their linear coefficient α; for an isotropic solid the volume coefficient is $\beta \approx 3\alpha$. Liquids and gases have only a volume coefficient β.

Solid	α (10^{-6} K^{-1})	Liquid or gas	β (10^{-4} K^{-1})
Aluminum	24	Ethanol	11.2
Brass	19	Gasoline	9.5
Copper	17	Glycerin	4.9
Steel	12	Water (20°C)	2.1
Concrete	12	Mercury	1.8
Glass (ordinary)	9	Air (0°C)	36.7
Glass (Pyrex)	3.2		
Invar (iron–nickel alloy)	0.9		

Example 36.2: Gaps between rails

Steel rails 12.0 m long are laid on a cool day at $10^{\circ}\mathrm{C}$. The hottest the rails are expected to get in summer sun is $45^{\circ}\mathrm{C}$. How large a gap must be left between neighboring rails so that they just touch on the hottest day?

Solution

Each rail lengthens by

$$\Delta L = \alpha L_0 \, \Delta T = (12 \times 10^{-6}\,\mathrm{K}^{-1})(12.0\,\mathrm{m})(35\,\mathrm{K}) = 5.0 \times 10^{-3}\,\mathrm{m} = \boxed{5.0\,\mathrm{mm}}. \tag{36.2.1}$$

Each rail grows by half this amount at either end, and so does its neighbor, so a gap of 5.0 mm between rail ends just closes. The steady click of a train over rail joints is the sound of these gaps.

Area and Volume Expansion

If every length grows by the factor $(1 + \alpha\ \Delta T)$, an area grows by that factor squared and a volume by that factor cubed. Because $\alpha\ \Delta T$ is tiny (around 10^{-4} or less), its square and cube are negligible, and

$$(1 + \alpha\,\Delta T)^2 \approx 1 + 2\alpha\,\Delta T, \qquad (1 + \alpha\,\Delta T)^3 \approx 1 + 3\alpha\,\Delta T. \tag{36.2.2}$$

So an area expands as $\Delta A = 2\alpha A_0\ \Delta T$, and a volume as

Key Equation: Volume thermal expansion

$$\Delta V = \beta V_0\,\Delta T, \qquad \beta \approx 3\alpha \text{ for a solid.} \tag{36.4}$$

Liquids have no fixed shape, so only their **coefficient of volume expansion** β means anything. Table 36.1 shows that liquids typically expand ten to a hundred times more than solids, which is the principle of the liquid-in-glass thermometer: the liquid swells faster than its glass tube, so it climbs the narrow bore.

A heated object expands like a photographic enlargement: every dimension scales by the same factor. That includes the empty spaces. A hole in a metal plate gets *larger* when the plate is heated, exactly as a circle drawn on the plate would, because the metal that would fill the hole would expand too. This is why a tight steel ring can be fitted onto a shaft by heating the ring first.

Example 36.3: An overflowing gas tank

A steel gas can is filled to the brim with 20.0 L of gasoline on a cool morning at $10^{\circ}\mathrm{C}$ and left in the sun, where both warm to $40^{\circ}\mathrm{C}$. How much gasoline spills out? Use Table 36.1.

Solution

Both the gasoline and the can expand, and the spill is the difference. The gasoline's volume grows by

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_0 \, \Delta T = (9.5 \times 10^{-4})(20.0\,\text{L})(30\,\text{K}) = 0.570\,\text{L}. \tag{36.2.3}$$

The inside of the can expands as if it were solid steel, with $\beta_{\text{steel}} = 3\alpha = 3.6 \times 10^{-5}\ \text{K}^{-1}$:

$$\Delta V_{\text{can}} = (3.6 \times 10^{-5})(20.0\,\text{L})(30\,\text{K}) = 0.0216\,\text{L}. \tag{36.2.4}$$

The spill is

$$\Delta V_{\text{gas}} - \Delta V_{\text{can}} = 0.570 - 0.022 \approx \boxed{0.55\,\text{L}}, \tag{36.2.5}$$

more than half a liter. The can's own expansion helps only a little, because steel expands about 25 times less than gasoline. That is why fuel cans and car tanks are never filled to the very top.

Checkpoint 36.2

A steel washer has a hole in its center. The washer is heated by $100\ \text{K}$. Does the hole get larger, get smaller, or stay the same size? Does the answer change if the washer is made of aluminum instead?

Answer

Different expansion coefficients can be put to work. A **bimetallic strip** is two thin strips of different metals, say brass and steel, bonded together along their length. When it is heated the brass lengthens more than the steel, and since the two are bonded, the strip must curl. Old-fashioned thermostats used a coiled bimetallic strip to tip a switch; the same principle still trips the circuit breakers in many homes when a wire carries too much current and warms.

Checkpoint 36.3

A straight brass–steel bimetallic strip is heated. Toward which metal does it bend: is the brass on the inside of the curve or on the outside? What happens if the strip is instead cooled below the temperature at which it was straight?

Answer

Thermal Stress

If an object is prevented from expanding, it pushes on whatever holds it. Clamp a rod between two rigid walls and heat it. The rod "wants" to grow by $\Delta L = \alpha L_0 \ \Delta T$, so the walls must squeeze it back by exactly that amount. In Chapter 12 we found that compressing a rod by ΔL requires a stress $F/A = Y \ \Delta L/L_0$, where Y is Young's modulus. Substituting the thermal ΔL, the **thermal stress** is

$$\frac{F}{A} = Y \alpha \, \Delta T. \tag{36.5}$$

Remarkably, the length of the rod has dropped out: a short rod and a long one develop the same stress for the same temperature change.

Example 36.4: Continuous welded rail

Modern railways often weld their rails into continuous lengths kilometers long, with no gaps at all, and anchor them firmly to the ties. Suppose a steel rail is welded in place at 20°C and the sun heats it to 50°C. Find (a) the thermal stress in the rail and (b) the compressive force on a rail of cross-sectional area $7.7 \times 10^{-3}\ \text{m}^2$. For steel, $Y = 2.0 \times 10^{11}\ \text{Pa}$ and $\alpha = 12 \times 10^{-6}\ \text{K}^{-1}$.

Solution

(a) Stress. From equation (36.5),

$$\frac{F}{A} = Y\alpha\,\Delta T = (2.0\times10^{11})(12\times10^{-6})(30) = \boxed{7.2\times10^{7}\,\mathrm{Pa}} = 72\,\mathrm{MPa}. \tag{36.2.6}$$

(b) Force. The force is the stress times the area:

$$F = (7.2\times10^{7}\,\mathrm{Pa})(7.7\times10^{-3}\,\mathrm{m}^2) \approx \boxed{5.5\times10^{5}\,\mathrm{N}}, \tag{36.2.7}$$

about the weight of a 56 tonne load, pushing along each rail. The stress is well below the roughly 250 MPa at which ordinary steel starts to yield, so the rail is not crushed. The danger is buckling: a long, thin rail under that much compression can kink sideways, so the rails are laid at a carefully chosen "neutral" temperature and held down by heavy anchors along their whole length.

The Strange Case of Water

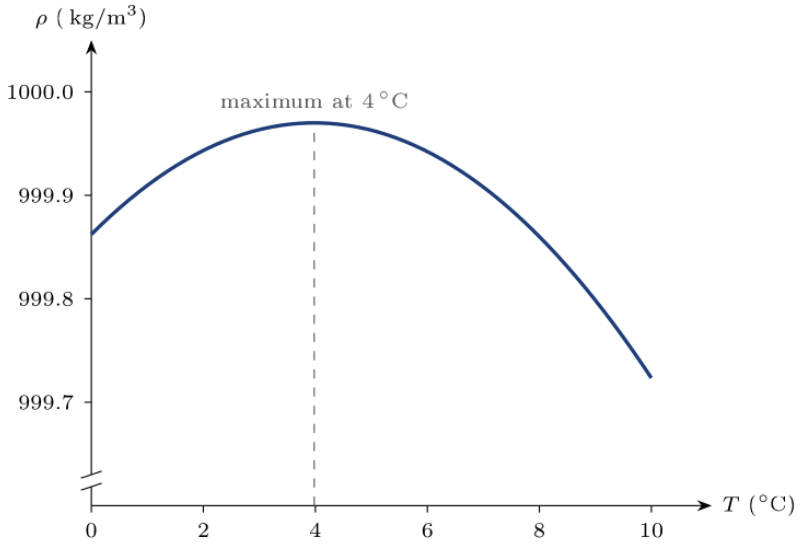


Figure 36.4. The density of liquid water near its freezing point. Unlike nearly every other liquid, water *contracts* as it warms from $0^{\circ}\mathrm{C}$ to $4^{\circ}\mathrm{C}$, reaching its greatest density at $4^{\circ}\mathrm{C}$, and only then begins to expand. (The vertical axis is greatly magnified: the whole variation shown is less than 0.03%.)

•Figure description

A graph of the density of liquid water against temperature from 0 to 10 degrees Celsius. The density rises from about 999.84 kilograms per cubic meter at 0 degrees to a maximum of 999.97 at 4 degrees, then falls again to about 999.70 at 10 degrees, tracing a gentle hump. A dashed vertical line marks the maximum at 4 degrees Celsius. The vertical axis starts at 999.6, not at zero, and a break mark on the axis shows this.

Water breaks the rule. Between $0^{\circ}\mathrm{C}$ and $4^{\circ}\mathrm{C}$ liquid water *shrinks* as it warms, so it is densest at $4^{\circ}\mathrm{C}$ (Figure 36.4). Ice, in turn, is about 8% *less* dense than the liquid (water swells by about 9% when it freezes), which is why ice floats (Chapter 14) and why a full bottle of water left in the freezer bursts. Both oddities come from the open, hexagonal network of hydrogen bonds that water molecules form as they approach freezing (hydrogen bonds are taken up in Chapter 42).

This peculiarity is why lakes freeze from the top down, and why fish survive the winter. As autumn air cools a lake, the cooled surface water becomes denser and sinks, stirring the lake, until the whole lake reaches $4^{\circ}\mathrm{C}$. Below that, cooling makes the surface water *less* dense, so it stays on top, cools further, and finally freezes into a floating lid. The ice insulates the water beneath, and the bottom of a deep lake stays near $4^{\circ}\mathrm{C}$ all winter.

Answers to Checkpoints

Checkpoint 36.2. The hole gets larger. Heating enlarges every dimension of the washer by the same factor, like a photographic

enlargement, and that includes the hole: the ring of metal around it grows in circumference, so the hole's diameter grows too. An aluminum washer behaves the same way, but its hole grows about twice as much, because aluminum's expansion coefficient is about twice steel's.

Checkpoint 36.3. Brass has the larger expansion coefficient, so when heated the brass side becomes longer than the steel side. A curved strip has a longer outside edge than inside edge, so the brass ends up on the *outside* of the curve and the strip bends toward the steel. When cooled, the brass shrinks more, so it ends up on the inside and the strip bends the other way, toward the brass.

This page titled 36.2: Thermal Expansion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.3: Ideal Gas Law

A solid or liquid has a definite volume that changes only a little with temperature. A gas has no volume of its own: it fills whatever container holds it, and its pressure, volume, and temperature are locked together. Experiments in the seventeenth to nineteenth centuries found three simple rules for a fixed amount of a dilute gas:

- **Boyle's law**: at constant temperature, the pressure is inversely proportional to the volume, $P \propto 1/V$. Squeeze a gas into half the space and its pressure doubles.
- **Charles's law**: at constant pressure, the volume is proportional to the Kelvin temperature, $V \propto T$. A balloon swells in the sun.
- **Gay-Lussac's law**: at constant volume, the pressure is proportional to the Kelvin temperature, $P \propto T$. This is the gas thermometer of Figure 36.1.

Combining the three, PV/T is constant for a fixed amount of gas, and the constant is proportional to how much gas there is.

To count the gas we use the **mole**, the SI unit for amount of substance. One mole contains exactly **Avogadro's number**

$$N_A = 6.022 \times 10^{23}\,\mathrm{mol}^{-1} \tag{36.3.1}$$

of molecules. The mass of one mole is the **molar mass** M; for any substance it is numerically equal to the mass of one molecule in atomic mass units, but expressed in grams per mole. Oxygen gas, $\mathrm{O_2}$, has $M = 32.0\ \mathrm{g/mol} = 0.0320\ \mathrm{kg/mol}$, and air, mostly nitrogen and oxygen, has an average $M = 28.9\ \mathrm{g/mol}$. A sample of mass m contains $n = m/M$ moles and $N = nN_A$ molecules.

With n moles of gas, the experimental laws combine into the **ideal gas law**:

Key Equation: Ideal gas law

$$PV = nRT = Nk_BT. \tag{36.6}$$

Here $R = 8.314\ \mathrm{J/(mol \cdot K)}$ is the **universal gas constant**, the same for every gas. The second form counts molecules instead of moles, and the constant that appears is **Boltzmann's constant**,

$$k_B = \frac{R}{N_A} = 1.381 \times 10^{-23}\,\mathrm{J/K}, \tag{36.3.2}$$

the gas constant per molecule. *In the ideal gas law T must be in kelvins and P must be the absolute pressure*, not the gauge pressure of Chapter 14. Using degrees Celsius would make the volume of a gas at $0^\circ\mathrm{C}$ equal to zero, which is absurd.

No real gas is exactly ideal. The law assumes that the molecules take up a negligible fraction of the volume and exert no forces on each other except when they collide. Both assumptions hold well for a gas at low density and at temperatures well above the point where it would condense. Air at room conditions obeys equation (36.6) to better than one part in a thousand.

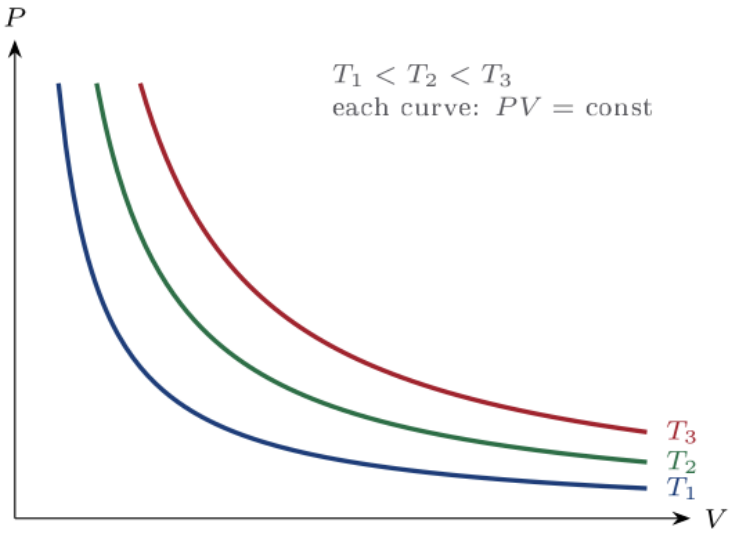


Figure 36.5. Isotherms of an ideal gas. At a fixed temperature, $PV = nRT$ is a constant, so the possible states lie on a hyperbola. Hotter isotherms lie farther from the origin. We will use these curves again in Chapters 37 and 38, where gases are taken from one state to another.

•Figure description

A pressure-versus-volume graph showing three curved lines called isotherms, each a hyperbola along which P times V is constant. The lowest curve is labeled T sub 1, the middle one T sub 2 and the highest one T sub 3, with T sub 1 less than T sub 2 less than T sub 3. Higher temperatures give curves farther from the origin.

On a graph of pressure against volume, called a P–V **diagram**, the states of a gas at one temperature lie on a hyperbola, $P = nRT/V$, called an **isotherm** (Figure 36.5). When a fixed amount of gas goes from one state to another, nR does not change, so

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}. \tag{36.7}$$

This form is handy whenever a gas changes state, because the quantities that do not change simply cancel.

A useful benchmark: one mole of any ideal gas at 0°C and one atmosphere (1.013×10^5 Pa), conditions called **standard temperature and pressure** (STP), occupies

$$V = \frac{nRT}{P} = \frac{(1)(8.314)(273.15)}{1.013 \times 10^5} = 0.0224\,\text{m}^3 = 22.4\,\text{L}, \tag{36.3.3}$$

about the volume of a large suitcase, whatever the gas.

Example 36.5: Molecules in a breath

You take a breath of 0.50 L of air, which warms to body temperature, 37°C, at a pressure of 1.01×10^5 Pa. How many molecules did you inhale?

Solution

Use the molecule-counting form of the ideal gas law, with everything in SI units: $V = 5.0 \times 10^{-4}$ m^3 and $T = 37 + 273 = 310$ K.

$$N = \frac{PV}{k_BT} = \frac{(1.01 \times 10^5)(5.0 \times 10^{-4})}{(1.381 \times 10^{-23})(310)} \approx \boxed{1.2 \times 10^{22}\,\text{molecules}}. \tag{36.3.4}$$

That is about ten thousand billion billion molecules in a single breath. It is so many that each breath you take almost certainly contains a few molecules that were once breathed by any historical figure you care to name, since the air of the entire atmosphere has had centuries to mix.

Example 36.6: Tire pressure on a cold morning

A car tire is inflated to a gauge pressure of 2.40×10^5 Pa (about 35 pounds per square inch) on a warm afternoon at 25°C. Overnight the tire cools to -5°C. Assuming the tire's volume does not change, what does the gauge read in the morning? Atmospheric pressure is 1.01×10^5 Pa.

Solution

The ideal gas law needs absolute pressures and Kelvin temperatures. The absolute pressure in the afternoon is $P_1 = 2.40 \times 10^5 + 1.01 \times 10^5 = 3.41 \times 10^5$ Pa, and the temperatures are $T_1 = 298$ K and $T_2 = 268$ K. With V and n fixed, equation (36.7) reduces to $P_1/T_1 = P_2/T_2$:

$$P_2 = P_1 \frac{T_2}{T_1} = (3.41 \times 10^5)\frac{268}{298} = 3.07 \times 10^5\,\text{Pa}. \tag{36.3.5}$$

The gauge reads the excess over atmospheric pressure:

$$P_{2,\text{gauge}} = 3.07 \times 10^5 - 1.01 \times 10^5 = \boxed{2.06 \times 10^5\,\text{Pa}}, \tag{36.3.6}$$

a drop of about 14%, which is why the low-tire-pressure light often comes on during the first cold snap of autumn. Had we carelessly used the gauge pressure and Celsius temperatures, we would have predicted $P_2 = 2.40 \times 10^5 \times (-5/25)$: a negative pressure, which is nonsense.

Example 36.7: A rising bubble

A scuba diver at a depth of 20.0 m in the sea, where the water is at 5°C, exhales a bubble. The bubble rises to the surface, where the water is at 20°C. By what factor has the bubble's volume increased? Take the density of seawater to be 1025 kg/m^3 and atmospheric pressure to be 1.013×10^5 Pa.

Solution

The pressure at depth is atmospheric plus the weight of the water above (Chapter 14):

$$P_1 = P_{\text{atm}} + \rho g h = 1.013 \times 10^5 + (1025)(9.80)(20.0) = 3.02 \times 10^5\,\text{Pa}. \tag{36.3.7}$$

The gas in the bubble keeps the same n, so from equation (36.7),

$$\frac{V_2}{V_1} = \frac{P_1}{P_2}\,\frac{T_2}{T_1} = \frac{3.02 \times 10^5}{1.013 \times 10^5} \cdot \frac{293}{278} = (2.98)(1.054) \approx \boxed{3.1}. \tag{36.3.8}$$

The bubble more than triples in volume, mostly because of the drop in pressure; the warming adds about 5%. The same expansion inside a diver's lungs is why a diver breathing compressed air must never hold the breath while ascending.

Checkpoint 36.4

A sealed, rigid container of gas is heated so that its Kelvin temperature doubles. By what factor does the pressure change? Would the answer be the same if the Celsius temperature doubled, from 20°C to 40°C?

Answer

Answers to Checkpoints

Checkpoint 36.4. At constant V and n, $PV = nRT$ gives $P \propto T$, so doubling the Kelvin temperature doubles the pressure. Doubling the Celsius temperature is not the same thing: 20°C to 40°C is 293 K to 313 K, so the pressure rises by only $313/293 = 1.07$, about 7%.

This page titled 36.3: Ideal Gas Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.4: Kinetic Theory of Gases

The ideal gas law is an experimental fact. The **kinetic theory of gases** explains *why* it holds, by treating a gas as an enormous number of tiny molecules in ceaseless random motion. The model makes four assumptions:

1. The gas consists of a very large number N of identical molecules, each of mass m, and the molecules themselves take up a negligible fraction of the container's volume.
2. The molecules move in random directions with a range of speeds, and obey Newton's laws.
3. The molecules exert no forces on each other or on the walls except during collisions, which are brief.
4. Collisions with the walls are elastic.

These are exactly the conditions under which a real gas behaves ideally.

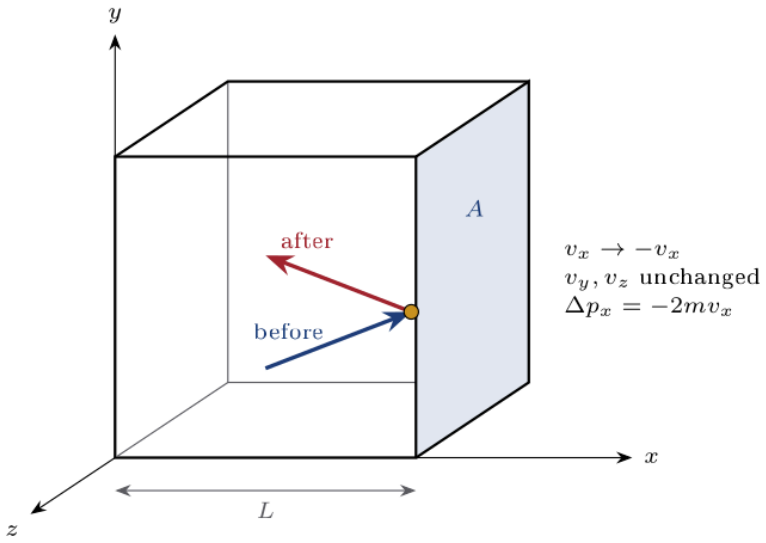


Figure 36.6. A molecule in a cubical box of edge L bounces elastically off the shaded wall of area $A = L^2$. Only the component of its velocity perpendicular to the wall reverses, so each collision changes its momentum by $-2mv_x$, and the wall receives an impulse of $+2mv_x$.

•Figure description

A cubical box of edge length L drawn in perspective, with x, y and z axes along three of its edges. A molecule, drawn as a small dot inside the box, moves toward the shaded right-hand wall, which has area A equals L squared. Its velocity before hitting the wall is drawn as an arrow with x component v sub x pointing toward the wall. After the elastic collision, a second arrow shows the velocity with its x component reversed to minus v sub x, while the y and z components are unchanged. The change in the molecule's momentum is labeled minus 2 m v sub x.

Pressure from molecular collisions. Put the gas in a cubical box of edge L and follow one molecule as it strikes the right-hand wall, which has area $A = L^2$ (Figure 36.6). In the elastic collision, the molecule's x component of velocity reverses from v_x to $-v_x$, while its other components are unchanged. Its momentum changes by $-2mv_x$, so by Newton's third law it gives the wall an impulse of $+2mv_x$. The molecule must then cross the box and back, a distance $2L$, before it hits the same wall again; that takes a time $\Delta t = 2L/v_x$. Averaged over many round trips, this one molecule pushes on the wall with an average force

$$\overline{F}_1 = \frac{\text{impulse per hit}}{\text{time between hits}} = \frac{2mv_x}{2L/v_x} = \frac{mv_x^2}{L}. \tag{36.4.1}$$

(Collisions with other molecules do not change this result, because on average they only swap momentum among molecules.) Adding up the forces from all N molecules gives $\overline{F} = (m/L)\sum v_x^2 = (Nm/L)\ \overline{v_x^2}$, where $\overline{v_x^2}$ is the average of v_x^2 over all the molecules.

The motion is random, so no direction is special, and the averages of v_x^2, v_y^2, and v_z^2 must be equal. Since $v^2 = v_x^2 + v_y^2 + v_z^2$, each one is a third of the average of v^2: $\overline{v_x^2} = \frac{1}{3}\overline{v^2}$. The pressure, force divided by area L^2, is therefore

$$P = \frac{\overline{F}}{L^2} = \frac{Nm\overline{v^2}}{3L^3} = \frac{1}{3}\frac{N}{V}m\overline{v^2} = \frac{2}{3}\frac{N}{V}\left(\tfrac{1}{2}m\overline{v^2}\right). \tag{36.8}$$

The pressure of a gas is proportional to the number of molecules per unit volume and to their average translational kinetic energy, $\overline{K} = \frac{1}{2}m\overline{v^2}$. Both make sense: more molecules mean more hits, and faster molecules hit harder and more often.

The meaning of temperature. Rewrite equation (36.8) as $PV = \frac{2}{3}N\overline{K}$ and compare it with the ideal gas law, $PV = Nk_BT$. The two agree only if

Key Equation: Average molecular kinetic energy

$$\overline{K} = \tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}k_BT. \tag{36.9}$$

This is one of the great results of physics. *The Kelvin temperature of a gas is a direct measure of the average translational kinetic energy of its molecules.* It does not depend on the kind of molecule, the pressure, or the volume. At room temperature, $300\ \text{K}$, every gas molecule has on average $\frac{3}{2}(1.381\times10^{-23})(300) = 6.2\times10^{-21}\ \text{J}$ of translational kinetic energy, about $0.04\ \text{eV}$. The quantity k_BT, about $4\times10^{-21}\ \text{J}$ or $\frac{1}{40}\ \text{eV}$ at room temperature, is the yardstick of thermal energy at the molecular scale; we used it in Chapters 16 and 23 to judge whether thermal jostling could scramble the alignment of molecular dipoles.

Taking the square root of $\overline{v^2}$ gives a typical molecular speed, the **root-mean-square speed**:

Key Equation: Root-mean-square speed

$$v_{\text{rms}} = \sqrt{\overline{v^2}} = \sqrt{\frac{3k_BT}{m}} = \sqrt{\frac{3RT}{M}}. \tag{36.10}$$

The second form uses the molar mass M (in kg/mol) instead of the mass of a molecule; the two are related by $M = N_Am$. At a given temperature lighter molecules move faster, because they must carry the same average kinetic energy with less mass.

Example 36.8: How fast are the molecules of the air?

Find the rms speed of (a) nitrogen molecules ($M = 28.0\ \text{g/mol}$) and (b) helium atoms ($M = 4.00\ \text{g/mol}$) at 20°C.

Solution

(a) Nitrogen. With $T = 293\ \text{K}$ and $M = 0.0280\ \text{kg/mol}$,

$$v_{\text{rms}} = \sqrt{\frac{3RT}{M}} = \sqrt{\frac{3(8.314)(293)}{0.0280}} = \sqrt{2.61\times10^5\,\text{m}^2/\text{s}^2} = \boxed{511\,\text{m/s}}. \tag{36.4.2}$$

(b) Helium. Only M changes, so the speed scales as $1/\sqrt{M}$:

$$v_{\text{rms,He}} = (511\,\text{m/s})\sqrt{\frac{28.0}{4.00}} = \boxed{1350\,\text{m/s}}. \tag{36.4.3}$$

Both are faster than sound in air ($343\ \text{m/s}$). That is no coincidence: sound is a pressure disturbance passed from molecule to molecule, so it cannot travel faster than the molecules that carry it, and its speed is a sizable fraction of v_{rms}. Sound in helium is correspondingly faster, which is why a voice sounds high-pitched after a breath of helium: the resonant frequencies of the vocal tract go up.

For a **monatomic** gas such as helium, neon, or argon, whose molecules are single atoms, translational kinetic energy is the only energy the molecules have (the atoms do not attract each other, so there is no potential energy). The **internal energy** of the gas, the total energy of all its molecules, is then N times equation (36.9):

$$E_{\text{int}} = \tfrac{3}{2}Nk_BT = \tfrac{3}{2}nRT \qquad (\text{monatomic ideal gas}). \tag{36.11}$$

The internal energy of an ideal gas depends only on its temperature, not on its pressure or volume. This fact will do a great deal

of work in the next chapter.

Checkpoint 36.5

(a) The Kelvin temperature of a gas is doubled. By what factor does v_{rms} change? (b) Hydrogen ($M = 2\ \text{g/mol}$) and oxygen ($M = 32\ \text{g/mol}$) are at the same temperature. Which molecules move faster, and by what factor?

Answer

Equipartition of Energy

A molecule of nitrogen or oxygen is two atoms bound together, shaped like a tiny dumbbell. Besides moving from place to place, it can tumble end over end, so it can store rotational kinetic energy as well. How is the energy shared between these motions?

The answer is the **equipartition theorem**, which follows from statistical mechanics: in thermal equilibrium, every independent way a molecule can store energy that depends on the square of a coordinate or a velocity (each **degree of freedom**) holds on average $\frac{1}{2}k_BT$. Translation has three degrees of freedom, $\frac{1}{2}mv_x^2$, $\frac{1}{2}mv_y^2$, and $\frac{1}{2}mv_z^2$, which is where the $\frac{3}{2}k_BT$ of equation (36.9) comes from. A diatomic molecule can also rotate about two axes perpendicular to its length, adding two more degrees of freedom. (Rotation about the axis through both atoms carries essentially no energy, because nearly all the mass sits on that axis.) With f degrees of freedom, the internal energy of an ideal gas is

$$E_{\text{int}} = \frac{f}{2}\,nRT, \qquad f = \begin{cases} 3 & \text{monatomic,} \\ 5 & \text{diatomic, near room temperature.} \end{cases} \tag{36.12}$$

The two atoms of a diatomic molecule can also vibrate along the bond, like masses on a spring, which would add two more degrees of freedom (kinetic and potential energy of the vibration). At room temperature, however, the vibrations are almost never excited. Vibrational energy is quantized (Chapter 42), and the smallest vibrational step of a nitrogen molecule is about ten times k_BT, so collisions are almost never energetic enough to start one. This “freezing out” of degrees of freedom was one of the first clues that classical physics fails at the molecular scale, and we will see in Chapter 37 how it shows up in the measured heat capacities of gases.

Checkpoint 36.6

Equal numbers of moles of helium and nitrogen are at the same temperature. (a) Which gas has the greater average *translational* kinetic energy per molecule? (b) Which has the greater internal energy? Explain both answers.

Answer

The Distribution of Molecular Speeds

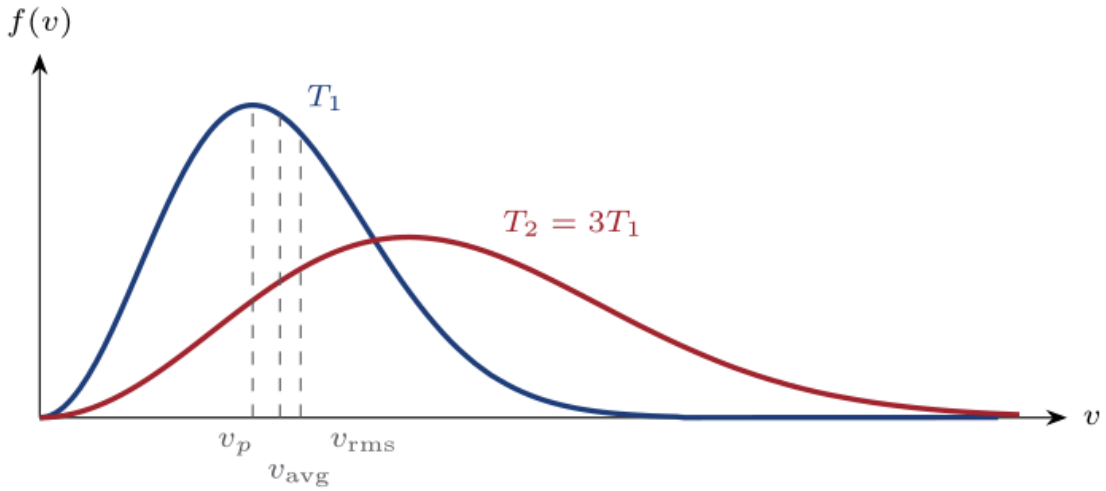


Figure 36.7. The Maxwell distribution of molecular speeds for the same gas at two temperatures. The area under a curve between two speeds is the fraction of molecules with speeds in that range, so each curve has total area 1. At the higher temperature the distribution shifts to higher speeds and spreads out. For the lower-temperature curve, dashed lines mark the most probable speed v_p, the average speed v_{avg}, and the rms speed v_{rms}.

•Figure description

A graph of the Maxwell speed distribution, the fraction of molecules per unit speed interval, against molecular speed v. Two curves are shown. The blue curve, for temperature T sub 1, rises from zero, peaks fairly sharply and then falls off with a long tail toward high speeds. Three dashed vertical lines under its peak region mark, from left to right, the most probable speed v sub p at the peak, the average speed v sub avg slightly to its right, and the rms speed v sub rms slightly farther right. The red curve, for a higher temperature T sub 2 equal to 3 T sub 1, is lower and broader and peaks at a higher speed. Both curves enclose the same total area.

The molecules of a gas do not all move at v_{rms}. Collisions constantly speed some up and slow others down, so at any instant there is a spread of speeds. In 1859 James Clerk Maxwell worked out its shape, the **Maxwell speed distribution** shown in Figure 36.7:

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 \, e^{-mv^2/2k_B T}. \tag{36.13}$$

The product $f(v)\ dv$ is the fraction of the molecules whose speeds lie between v and $v + dv$. The factor v^2 makes the curve start at zero (very few molecules are nearly at rest), and the exponential, a Boltzmann factor $e^{-K/k_B T}$, makes it fall off at high speeds. The fall-off is gradual, so the curve has a long tail: a small but important fraction of the molecules move several times faster than average.

The distribution has three natural “typical” speeds. The peak is the **most probable speed** v_p, the average of the speeds is v_{avg}, and the rms speed is v_{rms}. Integrating equation (36.13) gives

$$v_p = \sqrt{\frac{2k_B T}{m}}, \qquad v_{\text{avg}} = \sqrt{\frac{8k_B T}{\pi m}}, \qquad v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}}, \tag{36.14}$$

in the ratio $1 : 1.13 : 1.22$. They differ by only about 20%, so for most estimates any of them will do.

The high-speed tail matters far more than its size suggests. When water evaporates from a puddle or from your skin, the molecules that escape are the fastest ones in the tail, energetic enough to break free of their neighbors. Losing its fastest molecules lowers the average kinetic energy of the liquid left behind, so evaporation cools. The same tail lets a planet’s atmosphere leak away over billions of years, a light gas faster than a heavy one, as the last problem of this chapter explores. And the rates of chemical reactions, which need molecules colliding with more than some threshold energy, climb steeply with temperature because the tail grows so quickly.

Mean Free Path

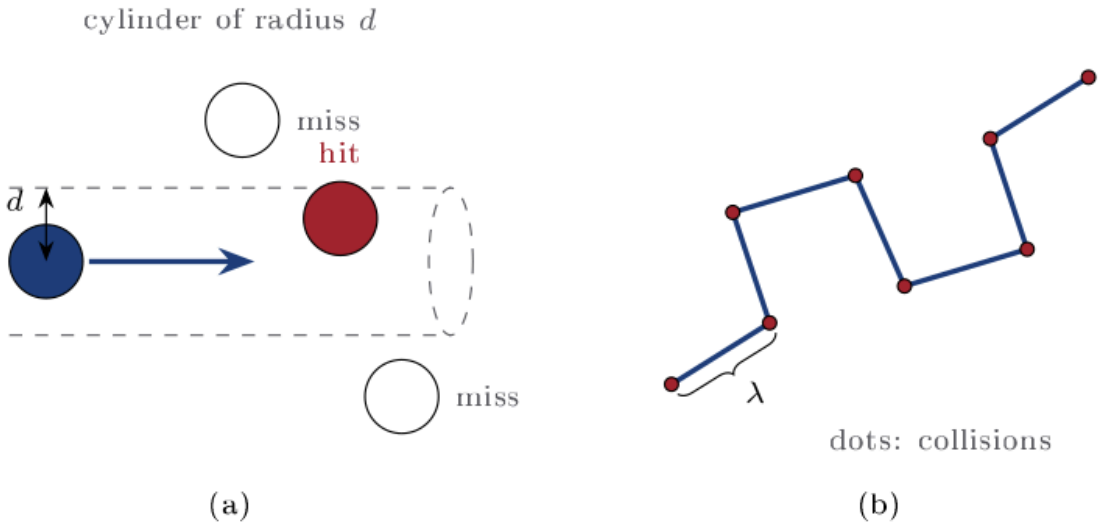


Figure 36.8. (a) A molecule of diameter d strikes any other molecule whose center lies within a distance d of its path, so it sweeps out a collision cylinder of cross-sectional area πd^2. (b) Its path through the gas is a random walk of straight steps between collisions, of average length λ, the mean free path.

•Figure description

Two diagrams side by side. On the left, a single molecule of diameter d moves to the right, and a dashed cylinder of radius d is drawn around its path: any other molecule whose center lies inside this cylinder will be struck. A few other molecules are shown: the one whose center lies inside the cylinder is labeled hit, and two outside it are labeled miss. On the right, a zigzag line shows the path of one molecule through a gas: a random walk of short straight steps, each ending at a collision. One step is labeled lambda, the mean free path.

We can now answer the question of the slow perfume. Real molecules are not points: they have a diameter d, about 3.7×10^{-10} m for the molecules of air, and they collide with each other. Two molecules collide if their centers come within a distance d of each other. As a molecule moves a distance ℓ, it sweeps out a cylinder of radius d and volume $\pi d^2 \ell$, and it hits every molecule whose center lies inside (Figure 36.8a). If there are N/V molecules per unit volume, the number of collisions in that distance is $(N/V)\pi d^2 \ell$. The average distance between collisions, the **mean free path** λ, is the ℓ that makes this number equal to one. Allowing for the motion of the other molecules adds a factor of $\sqrt{2}$:

Key Equation: Mean free path

$$\lambda = \frac{1}{\sqrt{2}\,\pi d^2\,(N/V)}. \tag{36.15}$$

The mean free path is shorter in a denser gas and for fatter molecules, as it should be.

Example 36.9: The slow perfume

Find (a) the mean free path of an air molecule ($d = 3.7 \times 10^{-10}$ m) at 20°C and atmospheric pressure, 1.01×10^5 Pa, and (b) the number of collisions it makes per second, taking its speed to be the average speed of nitrogen, $v_{\text{avg}} = 470$ m/s.

Solution

(a) Mean free path. First find the number of molecules per unit volume from the ideal gas law:

$$\frac{N}{V} = \frac{P}{k_B T} = \frac{1.01 \times 10^5}{(1.381 \times 10^{-23})(293)} = 2.50 \times 10^{25}\,\text{m}^{-3}. \tag{36.4.4}$$

Then

$$\lambda = \frac{1}{\sqrt{2}\,\pi(3.7\times10^{-10})^2(2.50\times10^{25})} = 6.6\times10^{-8}\,\text{m} \approx \boxed{66\,\text{nm}}, \tag{36.4.5}$$

about two hundred molecular diameters.

(b) Collision rate. The time between collisions is λ/v_{avg}, so the collision rate is

$$\frac{v_{\text{avg}}}{\lambda} = \frac{470}{6.6\times10^{-8}} \approx \boxed{7\times10^{9}\,\text{collisions/s}}. \tag{36.4.6}$$

A molecule of the air moves fast, but it changes direction seven billion times a second. Its path is a random walk, and a random walk covers ground slowly: after N steps of length λ the molecule is typically only about $\lambda\sqrt{N}$ from where it started, not $N\lambda$. Getting a perfume molecule across a room by pure diffusion would take days (the last problem of this chapter works out how many). The scent you notice in a minute or so was carried by air currents, gentle drafts and the convection from warm bodies and lamps, not by molecular motion alone.

The mean free path grows as the pressure falls. In the near-perfect vacuum of a particle accelerator or an electron microscope, λ is kilometers long, so the beam particles almost never hit a gas molecule on their way; in the thin air at the edge of space, a molecule can travel a long way between collisions, which is part of why fast molecules there can escape the planet altogether.

Answers to Checkpoints

Checkpoint 36.5. (a) $v_{\text{rms}} = \sqrt{3RT/M}$ is proportional to $\sqrt{T}$, so doubling the Kelvin temperature multiplies it by $\sqrt{2} \approx 1.41$. (b) At the same temperature, $v_{\text{rms}} \propto 1/\sqrt{M}$, so hydrogen is faster by $\sqrt{32/2} = \sqrt{16} = 4$. Both gases have the same average kinetic energy, $\frac{3}{2}k_BT$; the lighter molecules simply need a higher speed to carry it.

Checkpoint 36.6. (a) Neither: the average translational kinetic energy is $\frac{3}{2}k_BT$ for every gas at the same temperature. (b) The nitrogen. Its diatomic molecules also store rotational energy, two more degrees of freedom, so $E_{\text{int}} = \frac{5}{2}nRT$ for nitrogen against $\frac{3}{2}nRT$ for helium.

This page titled 36.4: Kinetic Theory of Gases is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.S: Temperature and Kinetic Theory of Gases (Summary)

Summary

- **Thermal equilibrium**: objects in contact stop exchanging energy when they reach the same **temperature**. The **zeroth law** (if A and B are each in equilibrium with C, they are in equilibrium with each other) is what makes thermometers meaningful.
- The pressure of a constant-volume gas extrapolates to zero at **absolute zero**, -273.15°C. Convert with $T_C = \frac{5}{9}(T_F - 32)$ and $T = T_C + 273.15$; a difference ΔT is the same in °C and K.
- **Thermal expansion**: $\Delta L = \alpha L_0\ \Delta T$, $\Delta A = 2\alpha A_0\ \Delta T$, $\Delta V = \beta V_0\ \Delta T$ with $\beta \approx 3\alpha$ for a solid. Holes expand like the material around them. A constrained rod develops a **thermal stress** $F/A = Y\alpha\ \Delta T$. Water is densest at 4°C.
- **Ideal gas law**: $PV = nRT = Nk_BT$, with T in kelvins and P absolute; $R = 8.314\ \text{J}/(\text{mol}\cdot\text{K})$, $k_B = R/N_A = 1.381 \times 10^{-23}\ \text{J/K}$. For a fixed amount of gas, $P_1V_1/T_1 = P_2V_2/T_2$. One mole occupies $22.4\ \text{L}$ at STP.
- **Kinetic theory**: pressure comes from molecular impacts, $P = \frac{2}{3}(N/V)\overline{K}$. Comparing with the ideal gas law, $\overline{K} = \frac{3}{2}k_BT$: temperature measures average translational kinetic energy. $v_{\text{rms}} = \sqrt{3k_BT/m} = \sqrt{3RT/M}$.
- **Equipartition**: each degree of freedom holds $\frac{1}{2}k_BT$ on average, so $E_{\text{int}} = \frac{f}{2}nRT$, with $f = 3$ for a monatomic gas and $f = 5$ for a diatomic gas near room temperature.
- The **Maxwell distribution** gives the spread of molecular speeds, with $v_p : v_{\text{avg}} : v_{\text{rms}} = 1 : 1.13 : 1.22$ and a long high-speed tail. The **mean free path** between collisions is $\lambda = 1/[\sqrt{2}\ \pi d^2(N/V)]$, about $66\ \text{nm}$ in air.

Key Equations

Temperature conversions

$$\begin{aligned} T_C &= \tfrac{5}{9}(T_F - 32) \\ T &= T_C + 273.15 \end{aligned} \tag{36.S.1}$$

Linear and volume thermal expansion

$$\Delta L = \alpha L_0\,\Delta T, \qquad \Delta V = \beta V_0\,\Delta T, \qquad \beta \approx 3\alpha \tag{36.S.2}$$

Thermal stress

$$\frac{F}{A} = Y\alpha\,\Delta T \tag{36.S.3}$$

Ideal gas law

$$PV = nRT = Nk_BT \qquad (T \text{ in kelvins},\ P \text{ absolute}) \tag{36.S.4}$$

Fixed amount of gas

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2} \tag{36.S.5}$$

Pressure from kinetic theory

$$P = \frac{2}{3}\,\frac{N}{V}\left(\tfrac{1}{2}m\overline{v^2}\right) \tag{36.S.6}$$

Average molecular kinetic energy

$$\overline{K} = \tfrac{3}{2} k_B T \tag{36.S.7}$$

Root-mean-square speed

$$v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}} = \sqrt{\frac{3RT}{M}} \tag{36.S.8}$$

Internal energy of an ideal gas

$$E_{\text{int}} = \frac{f}{2} nRT \qquad (f = 3 \text{ monatomic},\ f = 5 \text{ diatomic}) \tag{36.S.9}$$

Mean free path

$$\lambda = \frac{1}{\sqrt{2}\,\pi d^2\,(N/V)} \tag{36.S.10}$$

This page titled 36.S: Temperature and Kinetic Theory of Gases (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

36.E: Temperature and Kinetic Theory of Gases (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $R = 8.314\ \text{J/(mol·K)}$, $k_B = 1.381 \times 10^{-23}\ \text{J/K}$, $N_A = 6.022 \times 10^{23}\ \text{mol}^{-1}$, $P_{\text{atm}} = 1.01 \times 10^5\ \text{Pa}$, the expansion coefficients of Table 36.1, and other constants as given.

Section 36.1 — Temperature and Temperature Scales

36.1 ★ Convert: (a) 98.6°F (body temperature) to Celsius and Kelvin; (b) 25°C to Fahrenheit and Kelvin; (c) $0\ \text{K}$ to Celsius.

36.2 ★ *[Conceptual.]* Explain why a temperature *change* of 15°C is the same as a change of $15\ \text{K}$, but a temperature *reading* of 15°C is not $15\ \text{K}$.

36.3 ★ (a) The lowest air temperature ever recorded at Earth's surface is -89.2°C, at Vostok Station in Antarctica. Express it in degrees Fahrenheit and in kelvins. (b) At what temperature do the Fahrenheit and Celsius scales give the same numerical reading?

36.4 ★★ A constant-volume gas thermometer reads a pressure of $4.80 \times 10^4\ \text{Pa}$ when its bulb is in a mixture of ice and water at $273.15\ \text{K}$, and $6.45 \times 10^4\ \text{Pa}$ when the bulb is in a pot of boiling water on a mountain. Find the boiling temperature on the mountain in kelvins and in degrees Celsius, and explain why it is below 100°C.

Section 36.2 — Thermal Expansion

36.5 ★ A steel bridge span is $300\ \text{m}$ long at 10°C. How much longer is it on a hot day at 40°C?

36.6 ★ *[Conceptual.]* Casserole dishes and laboratory glassware are made of Pyrex rather than ordinary glass. Using Table 36.1, explain why Pyrex is much less likely to crack when hot food is put into a cold dish or a cold liquid is poured into a hot beaker.

36.7 ★★ An aluminum ring has an inner diameter of $4.980\ \text{cm}$ at 20°C. It must be slipped onto a steel shaft $5.000\ \text{cm}$ in diameter, also at 20°C. To what temperature must the ring be heated? (Once it cools, the ring grips the shaft tightly; this is called a shrink fit.)

36.8 ★★ A Pyrex flask ($\beta = 9.6 \times 10^{-6}\ \text{K}^{-1}$) is filled to the brim with $250.0\ \text{mL}$ of ethanol at 10°C. The flask and the ethanol are warmed together to 40°C. How much ethanol overflows?

36.9 ★★ A brass rod is clamped between two rigid walls at 20°C, just touching both without stress. It is then heated to 70°C. Find the thermal stress in the rod. For brass, $Y = 9.0 \times 10^{10}\ \text{Pa}$.

36.10 ★★ *[Real-world application: timekeeping.]* A pendulum clock with a brass pendulum rod keeps perfect time at 20°C. Treat it as a simple pendulum (Chapter 28), whose period is proportional to the square root of its length. (a) Show that a small temperature change ΔT changes the period by the fraction $\Delta T_{\text{per}}/T_{\text{per}} \approx \tfrac{1}{2}\alpha\ \Delta T$. (b) In a summer heat wave the clock is at 35°C. Does it run fast or slow, and by how many seconds per day? (c) Why were the pendulums of precision clocks made of Invar?

Section 36.3 — Ideal Gas Law

36.11 ★ A $2.0\ \text{mol}$ sample of ideal gas is held at $300\ \text{K}$ in a $0.025\ \text{m}^3$ container. Find its pressure.

36.12 ★ A room measures $4.0\ \text{m} \times 5.0\ \text{m} \times 2.5\ \text{m}$ and is filled with air at 20°C and atmospheric pressure. (a) How many moles of air does it hold? (b) What is the mass of the air, if $M = 28.9\ \text{g/mol}$?

36.13 ★ *[Conceptual.]* A sealed bag of potato chips, packed at sea level, puffs up noticeably during an airplane flight or a drive into the mountains. Explain why, using the ideal gas law.

36.14 ★★ A scuba tank of volume $12.0\ \text{L}$ is filled with air at an absolute pressure of $2.00 \times 10^7\ \text{Pa}$ and a temperature of 20°C

. (a) What volume would this air occupy at atmospheric pressure and the same temperature? (b) A diver at a depth of $10.0\ \text{m}$ in seawater ($\rho = 1025\ \text{kg/m}^3$) breathes $25\ \text{L}$ per minute of air at the surrounding water pressure. Ignoring the reserve that must stay in the tank, how long does the tank last? Assume the temperature stays at 20°C.

36.15 ★★ *[Real-world application: meteorology.]* A weather balloon is released at sea level, where the pressure is $1.01 \times 10^5\ \text{Pa}$ and the temperature is 15°C, with its helium filling a sphere of radius $1.00\ \text{m}$. It rises to an altitude of $30\ \text{km}$, where the pressure is $1.20 \times 10^3\ \text{Pa}$ and the temperature is -46°C. Assuming the balloon's skin stretches freely, so the helium stays at the outside pressure, find the balloon's radius at that altitude.

Section 36.4 — Kinetic Theory of Gases

36.16 ★ Find the average translational kinetic energy of a gas molecule at $300\ \text{K}$, in joules and in electron-volts ($1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}$).

36.17 ★★ Oxygen has a molar mass of $32.0\ \text{g/mol}$. At what temperature would the rms speed of oxygen molecules equal Earth's escape speed, $11.2\ \text{km/s}$?

36.18 ★★ (a) Find the internal energy of $2.0\ \text{mol}$ of helium gas at $300\ \text{K}$. (b) Find the internal energy of $2.0\ \text{mol}$ of nitrogen at the same temperature. (c) Account for the difference in terms of degrees of freedom.

36.19 ★★ For nitrogen ($M = 28.0\ \text{g/mol}$) at $300\ \text{K}$, calculate v_p, v_{avg}, and v_{rms}, and verify that they are in the ratio $1 : 1.13 : 1.22$.

36.20 ★★ The best laboratory vacuum systems reach a pressure of about $1.0 \times 10^{-8}\ \text{Pa}$. At this pressure and 20°C, find (a) the number of gas molecules per cubic centimeter and (b) the mean free path of a molecule with $d = 3.7 \times 10^{-10}\ \text{m}$. Compare the mean free path with the size of the vacuum chamber, a meter or so.

36.21 ★★ *[Conceptual.]* (a) Why does evaporation cool the liquid left behind? Use the Maxwell distribution in your answer. (b) Why does a fan make you feel cooler on a hot, dry day even though it does not lower the temperature of the air?

Challenge Problems

36.22 ★★★ *[Real-world application: atmospheric science — why Earth has no hydrogen atmosphere.]* A gas molecule escapes a planet's atmosphere if it is moving fast enough, upward, at an altitude where collisions are rare. Because the Maxwell distribution has a long high-speed tail, a useful rule of thumb is that a planet loses a gas over geological time if v_{rms} exceeds roughly one sixth of the escape speed. Take the exosphere temperature to be $1000\ \text{K}$ for Earth ($v_{\text{esc}} = 11.2\ \text{km/s}$) and $400\ \text{K}$ for the Moon ($v_{\text{esc}} = 2.38\ \text{km/s}$).

a. Find v_{rms} for H_2 ($M = 2.016\ \text{g/mol}$) and for N_2 ($M = 28.0\ \text{g/mol}$) at $1000\ \text{K}$.
b. Apply the rule of thumb to each gas at Earth. Which does Earth retain, and which has it lost?
c. Repeat for N_2 on the Moon at $400\ \text{K}$, and comment on why the Moon has essentially no atmosphere at all.
d. Titan, a moon of Saturn, has $v_{\text{esc}} = 2.64\ \text{km/s}$ (barely more than our Moon's) yet holds a thick nitrogen atmosphere. Its exosphere temperature is about $180\ \text{K}$. Show with a calculation why this is consistent with the rule of thumb, and identify which single variable does the most work in the comparison.

36.23 ★★★ *[Real-world application: the slow perfume.]* Reread the chapter opener and the example on the mean free path. A molecule in air at 20°C and atmospheric pressure has a mean free path of $66\ \text{nm}$ and makes about 7×10^9 collisions per second.

a. In a random walk of N steps of length λ, the typical distance from the starting point is $\lambda\sqrt{N}$. How many collisions does a molecule need to make to get $5.0\ \text{m}$ from where it started?
b. How long does that take, in days?
c. If the molecule had instead flown straight at $470\ \text{m/s}$, how long would the trip have taken? By what factor does the random walk slow it down?
d. At an altitude of $100\ \text{km}$ the air's density is about 2×10^6 times lower than at sea level. Estimate the mean free path there, and explain why the boundary between "air" and "space" is fuzzy rather than sharp.

This page titled 36.E: Temperature and Kinetic Theory of Gases (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37: HEAT AND THE FIRST LAW OF THERMODYNAMICS

Drive inland from the California coast on a summer afternoon and you can watch the thermometer climb. A morning that was a mild $18^{\circ}\mathrm{C}$ in Monterey becomes a baking $38^{\circ}\mathrm{C}$ by the time you reach Fresno, a little over two hours east. The two cities sit at almost exactly the same latitude and share the same sun; what they do not share is the ocean. The coast sits beside an enormous reservoir of water that warms and cools only sluggishly, holding the shoreline's temperature nearly constant day and night, summer and winter, while the dry inland air swings wildly between hot and cold. The ocean works like a huge savings account for heat: the sunny day's deposits and the cool night's withdrawals are tiny compared with what is already in the account, so its balance, the water's temperature, hardly moves. By the end of this chapter you will be able to say exactly why water behaves this way (it comes down to a single number, water's unusually large *specific heat*) and to calculate how much energy it takes to change the temperature of anything, from a bullet to a swimming pool.

Chapter 36 was about temperature and the state of matter. This chapter is about energy in transit. Energy can enter or leave a system in two ways: as **heat**, because of a temperature difference, and as **work**, because of a force acting through a distance. We begin with heat: how much it takes to warm something or to melt or boil it, and how to balance the books when hot and cold things are mixed. We then combine heat and work into the **first law of thermodynamics**, which is conservation of energy rewritten to include heat, and apply it to the gases of the last chapter as they are heated, compressed, and expanded. The chapter ends with the three ways heat travels from place to place. Chapter 38 takes up the deeper question of which way heat is allowed to flow.

Learning Objectives

- Distinguish heat, temperature, and internal energy, and convert between joules and calories.
- Compute the heat needed to change a substance's temperature or to change its phase.
- Solve calorimetry problems using conservation of energy.
- Calculate the work done by a gas as the area under its path on a P–V diagram, and explain why heat and work depend on the path.
- Apply the first law of thermodynamics, $\Delta E_{\text{int}} = Q + W$, to isobaric, isovolumetric, isothermal, and adiabatic processes, using the molar specific heats of an ideal gas.
- Calculate heat flow by conduction and radiation, and describe convection.

We will return to the coastal-climate question once we reach specific heat, and again in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

37.1: Heat
37.2: The First Law of Thermodynamics
37.3: Thermodynamic Processes
37.4: Heat Transfer Mechanisms
37.S: Heat and the First Law of Thermodynamics (Summary)
37.E: Heat and the First Law of Thermodynamics (Problems)

This page titled 37: Heat and the First Law of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.1: Heat

Heat is energy transferred between a system and its surroundings because of a temperature difference between them. We give it the symbol Q, and the crucial directional fact is that heat always flows *spontaneously from higher temperature to lower temperature*, never the other way on its own: a one-way street we will examine closely in Chapter 38.

It is worth being careful with words, because everyday language blurs three different ideas.

- **Temperature** measures the average kinetic energy of a substance's molecules (Chapter 36). It says nothing about how much matter there is.
- **Internal energy** E_{int} is the total energy of all the molecules: their kinetic energy of translation, rotation, and vibration, and the potential energy of the forces between them. It is a property of the system, like its mass or its volume.
- **Heat** is not something a system *has*. It is energy in transit, crossing the boundary because of a temperature difference. Once it has arrived, it is simply part of the internal energy.

A swimming pool and a cup of coffee can be at the same temperature, yet the pool has vastly more internal energy and can deliver far more heat, simply because it contains far more matter. And a hot object does not contain "heat" any more than a bank account contains "deposits": deposits are what came in, but the account holds money.

Heat is energy, so its SI unit is the joule. Two older units persist. The **calorie** (cal) is the heat needed to raise one gram of water by one Celsius degree, and the **Calorie** (Cal, capital C) used on food labels is actually a kilocalorie:

$$1\,\text{cal} = 4.186\,\text{J}, \qquad 1\ \text{Cal} = 1000\ \text{cal} = 4186\,\text{J}. \tag{37.1.1}$$

The candy bar that lists "250 Calories" is quoting a quarter of a million of the physicist's calories.

That heat and work are measured in the same unit was not obvious. Through the eighteenth century heat was thought to be an invisible fluid, "caloric," that flowed from hot bodies into cold ones. In the 1840s James Prescott Joule showed otherwise. In his best-known experiment, falling weights turned a paddle wheel in an insulated tank of water, and the stirred water warmed by exactly the amount that the same quantity of heat would have warmed it. Mechanical work and heat were interchangeable, at a fixed exchange rate of $4.186\ \text{J}$ per calorie, the **mechanical equivalent of heat**. Both are simply ways of transferring energy.

Example 37.1: Working off a candy bar

A $70\ \text{kg}$ hiker eats a 250-Calorie candy bar and wants to "burn it off" by climbing stairs. (a) If all of the food energy went into gravitational potential energy, how high would the hiker have to climb? (b) Human muscles convert food energy to mechanical work with an efficiency of only about 25%; the rest becomes internal energy and is shed as heat. How high must the hiker really climb?

Solution

(a) The food energy is $250 \times 4186\ \text{J} = 1.05 \times 10^{6}\ \text{J}$. Setting it equal to mgh,

$$h = \frac{E}{mg} = \frac{1.05 \times 10^{6}}{(70)(9.8)} \approx \boxed{1500\,\text{m}}. \tag{37.1.2}$$

(b) Only a quarter of the energy does mechanical work, so the climb is a quarter as high:

$$h = (0.25)(1500\,\text{m}) \approx \boxed{380\,\text{m}}, \tag{37.1.3}$$

about the height of the Empire State Building's roof. The other 75% warms the body, which is why a stair climb makes you sweat. Food energy is a great deal of energy by mechanical standards.

Temperature Changes

How much heat does it take to warm a given object by a given amount? The answer depends on the object's mass and on a property of its material called the **specific heat capacity** c, or simply specific heat: the heat required to raise the temperature of one kilogram of the substance by one kelvin. The relation is

Key Equation: Heat and temperature change

$$Q = mc\,\Delta T, \tag{37.1}$$

where m is the mass, c the specific heat (units $\mathrm{J/(kg\cdot K)}$), and $\Delta T = T_f - T_i$ the temperature change. A positive Q (heat in) raises the temperature, and a negative Q (heat out) lowers it. The product $C = mc$ is the **heat capacity** of a particular object, the heat needed to raise its temperature by one kelvin.

Table 37.1. Specific heats of some common substances near room temperature and atmospheric pressure.

Substance	c ($\mathrm{J/(kg\cdot K)}$)	Substance	c ($\mathrm{J/(kg\cdot K)}$)
Water (liquid)	4186	Granite	790
Human body (average)	3500	Iron	452
Ethanol	2430	Copper	387
Ice ($-10^\circ\mathrm{C}$)	2090	Silver	235
Steam ($110^\circ\mathrm{C}$)	2010	Mercury	140
Aluminum	900	Lead	128
Glass	840		

A large c means a substance is hard to heat up and hard to cool down: it stores a lot of energy per degree. Water is the standout in Table 37.1: its specific heat is several times that of most solids and nearly five times aluminum's. (The human body, mostly water, comes close.)

This single number answers the question we opened with. The ocean's enormous specific heat lets it absorb and release tremendous quantities of heat with only a small change in its own temperature, so the air above it, and the towns along the coast, stay mild. Inland, dry soil and rock (specific heats near $800\ \mathrm{J/(kg\cdot K)}$ or less) heat up and cool down quickly, and the temperature swings to extremes. The effect is magnified because sunlight penetrates many meters into clear water and the waves mix it, so a great depth of ocean shares each day's heat, while on land only the top few centimeters of soil warm up.

Checkpoint 37.1

Two identical pots, one holding $1\ \mathrm{kg}$ of water and the other $1\ \mathrm{kg}$ of aluminum, receive the same amount of heat. Which ends up hotter, and by roughly what factor is its temperature change larger? *(Use equation (37.1) with Table 37.1.)*

Answer

Example 37.2: Heating a bullet on impact

A silver bullet is fired at $200\ \mathrm{m/s}$ into a wooden wall and stops. Assume all of the bullet's kinetic energy becomes internal energy that stays in the bullet.

a. What is the temperature change of the silver bullet? (For silver, $c = 235\ \mathrm{J/(kg\cdot K)}$.)
b. If a lead bullet ($c = 128\ \mathrm{J/(kg\cdot K)}$) were used instead, would its temperature change be larger or smaller?

Solution

(a) The kinetic energy all goes into heating the bullet, $\frac{1}{2}mv^2 = mc\ \Delta T$. The mass cancels:

$$\Delta T = \frac{v^2}{2c} = \frac{(200\,\mathrm{m/s})^2}{2(235\,\mathrm{J/(kg\cdot K)})} = \frac{40000}{470} \approx \boxed{85\,\mathrm{K}}. \tag{37.1.4}$$

(b) Since $\Delta T = v^2/2c$, a *smaller* specific heat gives a *larger* temperature change. Lead's c is smaller than silver's, so the lead bullet heats up $\boxed{\text{more}}$: by $(200)^2/2(128) \approx 156$ K, nearly twice as much, which takes it a good part of the way toward lead's low melting point of 327°C.

Phase Changes

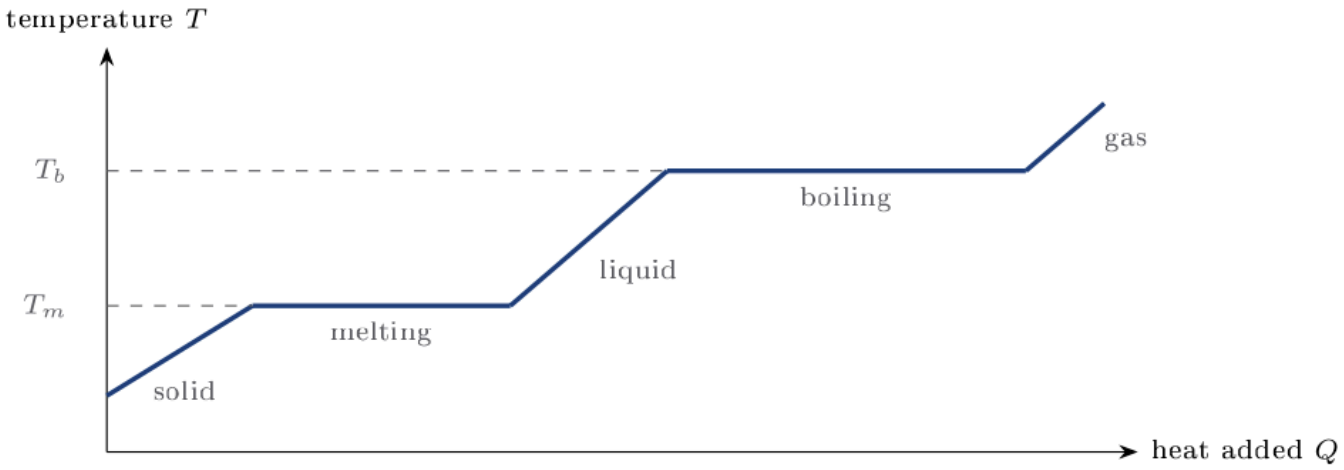


Figure 37.1. Heating a substance through its phase changes. Along the sloped segments the temperature rises ($Q = mc\ \Delta T$); along the flat plateaus the temperature holds steady while a phase change absorbs the heat ($Q = mL$). The melting plateau and the longer boiling plateau are the latent heats at work.

•Figure description

A graph of temperature against heat added as a substance is warmed from solid to gas. The curve rises in a straight sloped segment labeled solid, then runs flat along a plateau labeled melting at the melting temperature T sub m, then rises again through a segment labeled liquid, then runs flat along a longer plateau labeled boiling at the boiling temperature T sub b, and finally rises again as gas. On the sloped parts the heat raises the temperature; on the flat plateaus it drives a phase change instead, and the boiling plateau is far longer than the melting one.

Add heat to a block of ice and its temperature climbs, until it reaches 0°C. There the temperature *stops rising* even as you keep adding heat, while the ice melts to water; only once all the ice is gone does the temperature start to climb again (Figure 37.1). The heat poured in during melting does not raise the temperature, so it does not speed up the molecules; it goes into breaking the bonds that hold the solid together, raising the molecules' potential energy instead. This energy of a phase change is the **latent heat** (literally "hidden" heat), and the heat to change the phase of a mass m is

Key Equation: Heat of a phase change

$$Q = mL, \tag{37.2}$$

where L is the latent heat of the substance (units $\mathrm{J/kg}$). There are three kinds: the **latent heat of fusion** L_f for melting or freezing (solid $\leftrightarrow$ liquid), the **latent heat of vaporization** L_v for boiling or condensing (liquid $\leftrightarrow$ gas), and the **latent heat of sublimation** L_s for the direct solid $\leftrightarrow$ gas change. The same L applies in both directions: freezing a kilogram of water releases exactly the 3.33×10^5 J it took to melt it. When applying equation (37.2), take Q positive for melting and boiling (heat in) and negative for freezing and condensing (heat out).

Table 37.2. Melting and boiling points at atmospheric pressure, with latent heats of fusion and vaporization.

Substance	Melting point	L_f (J/kg)	Boiling point	L_v (J/kg)
Nitrogen	-210°C	2.55×10^4	-196°C	2.01×10^5

Substance	Melting point	L_f (J/kg)	Boiling point	L_v (J/kg)
Ethanol	-114°C	1.04×10^5	78°C	8.54×10^5
Mercury	-39°C	1.18×10^4	357°C	2.72×10^5
Water	0°C	3.33×10^5	100°C	2.26×10^6
Lead	327°C	2.45×10^4	1750°C	8.70×10^5
Aluminum	660°C	3.97×10^5	2450°C	1.14×10^7

For water the values are large (Table 37.2): $L_f = 3.33 \times 10^5$ J/kg and $L_v = 2.26 \times 10^6$ J/kg. Vaporizing water takes nearly seven times the energy of melting it, because boiling must pull the molecules completely apart while melting only loosens them. That is why steam burns are so severe (condensing on your skin, steam delivers its large latent heat) and why sweating cools you so effectively.

Example 37.3: From ice to steam

How much heat does it take to turn 1.00 kg of ice at -10°C into steam at 110°C? Which step takes the most?

Solution

Follow Figure 37.1 step by step, using Tables 37.1 and 37.2.

$$\begin{aligned}
\text{warm the ice, } -10 \to 0^\circ\text{C}: &\quad Q_1 = (1.00)(2090)(10) = 20900\,\text{J} \\
\text{melt the ice:} &\quad Q_2 = (1.00)(3.33 \times 10^5) = 333000\,\text{J} \\
\text{warm the water, } 0 \to 100^\circ\text{C}: &\quad Q_3 = (1.00)(4186)(100) = 418600\,\text{J} \\
\text{boil the water:} &\quad Q_4 = (1.00)(2.26 \times 10^6) = 2260000\,\text{J} \\
\text{warm the steam, } 100 \to 110^\circ\text{C}: &\quad Q_5 = (1.00)(2010)(10) = 20100\,\text{J}
\end{aligned} \tag{37.1.5}$$

The total is

$$Q = Q_1 + Q_2 + Q_3 + Q_4 + Q_5 = \boxed{3.05 \times 10^6\,\text{J}}. \tag{37.1.6}$$

Boiling alone takes $2.26/3.05 = 74\%$ of it, more than five times the heat needed to warm the water all the way from freezing to boiling. This is why a pot left on the stove takes so long to boil dry.

Calorimetry

When a hot object and a cold object are placed together and isolated from everything else, heat flows from hot to cold until they reach a common temperature. Energy is conserved, so the heat lost by the hot object equals the heat gained by the cold one. This bookkeeping, **calorimetry**, is the standard way to measure specific heats and latent heats. The rule is simply that the heat gains and losses sum to zero,

$$\sum Q = 0, \tag{37.3}$$

where each Q is computed with $Q = mc\,\Delta T$ (and $Q = mL$ for any phase change along the way), taking $\Delta T = T_{\text{final}} - T_{\text{initial}}$ so that objects that cool contribute negative Q and objects that warm contribute positive Q.

Example 37.4: Specific heat of a mystery liquid

A piece of glass at 85.0°C is covered by an equal mass of liquid at 43.0°C. The two reach equilibrium at 57.0°C. Ignoring the container and any loss to the surroundings, find the specific heat of the liquid. (For glass, $c = 840$ J/(kg · K).)

Solution

The glass cools and the liquid warms; the heat lost by the glass equals the heat gained by the liquid. With equal masses m, the m cancels:

$$\underbrace{mc_{\text{glass}}(85.0-57.0)}_{\text{lost by glass}} = \underbrace{mc_{\text{liq}}(57.0-43.0)}_{\text{gained by liquid}}. \tag{37.1.7}$$

Solving for the liquid's specific heat,

$$c_{\text{liq}} = c_{\text{glass}}\,\frac{85.0-57.0}{57.0-43.0} = (840)\frac{28.0}{14.0} = \boxed{1680\,\text{J/(kg}\cdot\text{K)}}. \tag{37.1.8}$$

The liquid changed temperature by only 14° while the glass changed by 28°, so for equal masses the liquid must store more energy per degree: consistent with its larger specific heat.

Example 37.5: Finding a metal's specific heat

A $4.2\ \text{kg}$ metal bar at 114°C is dropped into a $2.3\ \text{kg}$ aluminum calorimeter containing $3.9\ \text{kg}$ of water, both initially at 18°C. The system reaches equilibrium at 42°C. Find the specific heat of the metal. (Water $c = 4186$, aluminum $c = 900\ \text{J/(kg}\cdot\text{K)}$.)

Solution

The bar cools from 114° to 42° ($\Delta T = 72^\circ$); the water and the aluminum cup both warm from 18° to 42° ($\Delta T = 24^\circ$). Heat lost by the bar equals heat gained by water plus aluminum:

$$m_{\text{bar}}c_{\text{bar}}(72) = m_w c_w(24) + m_{\text{Al}}c_{\text{Al}}(24). \tag{37.1.9}$$

The right side is

$$(3.9)(4186)(24) + (2.3)(900)(24) = 391{,}810 + 49{,}680 = 441{,}490\,\text{J}. \tag{37.1.10}$$

The left side is $m_{\text{bar}}c_{\text{bar}}(72) = (4.2)(72)\ c_{\text{bar}} = 302.4\ c_{\text{bar}}$, so

$$c_{\text{bar}} = \frac{441{,}490}{302.4} \approx \boxed{1460\,\text{J/(kg}\cdot\text{K)}}. \tag{37.1.11}$$

That is a large value for a metal (most metals lie below $1000\ \text{J/(kg}\cdot\text{K)}$; beryllium, at about $1820\ \text{J/(kg}\cdot\text{K)}$, is a rare exception), but it does fall between water's and aluminum's: a sanity check that we have not lost a factor somewhere.

Example 37.6: Punch and ice: does it all melt?

At a party, $1.00\ \text{kg}$ of punch (treat it as water) sits at 24°C in a thermally isolated bowl. Someone adds $500\ \text{g}$ of ice at -10°C. Find the final temperature. (Ice $c = 2090$, water $c = 4186$, $L_f = 3.33\times10^5\ \text{J/kg}$.)

Solution

The question in the title is the key: we must not assume all the ice melts. Tally the energy demands and supply, taking 0°C as the meeting point.

Heat the punch can give up cooling from 24° to 0°:

$$Q_{\text{punch}} = mc_w\Delta T = (1.00)(4186)(24) = 100{,}500\,\text{J}. \tag{37.1.12}$$

Heat to warm the ice from -10° to 0°:

$$Q_{\text{warm}} = mc_{\text{ice}}\Delta T = (0.500)(2090)(10) = 10{,}450\,\text{J}. \tag{37.1.13}$$

Heat to melt all the ice:

$$Q_{\text{melt}} = mL_f = (0.500)(3.33\times10^5) = 166{,}500\,\text{J}. \tag{37.1.14}$$

After the punch warms the ice to 0° it has $100{,}500 - 10{,}450 = 90{,}050\ \mathrm{J}$ left, not enough to melt all $166{,}500\ \mathrm{J}$ worth of ice. So the ice does *not* all melt, and the system settles at

$$\boxed{T_f = 0^\circ\mathrm{C}}, \tag{37.1.15}$$

an ice–water slush. The leftover energy melts only $m = (90{,}050)/(3.33 \times 10^5) \approx 0.27\ \mathrm{kg}$ of the ice, leaving about $0.23\ \mathrm{kg}$ unmelted.

Answers to Checkpoints

Checkpoint 37.1. The aluminum. With the same Q and mass, $\Delta T = Q/(mc)$ is inversely proportional to the specific heat. Water's specific heat is about $4186\ \mathrm{J/(kg \cdot K)}$ and aluminum's about $900\ \mathrm{J/(kg \cdot K)}$, so the aluminum's temperature rises roughly $4186/900 \approx 4.7$ times as much.

This page titled 37.1: Heat is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.2: The First Law of Thermodynamics

We now broaden conservation of energy to include heat. Describe a gas by its **state variables**: pressure P, volume V, temperature T, and internal energy E_{int}, each a property of the system's current state. We can change that state by two kinds of energy transfer, the **transfer variables** heat Q and work W. Imagine a gas in a cylinder fitted with a piston: piling sand on the piston does *work* that compresses the gas, while a flame beneath supplies *heat* that can expand it. Both change the gas's internal energy, and the accounting is the **first law of thermodynamics**:

Key Equation: First law of thermodynamics

$$\Delta E_{\text{int}} = Q + W. \tag{37.4}$$

This is conservation of energy, nothing more: the change in a system's internal energy equals the heat added to it plus the work done *on* it. The sign convention matters. Here $Q > 0$ means heat flows *into* the system, and $W > 0$ means work is done *on* the system (for example, compressing it). When a gas instead *expands* and pushes on its surroundings, it does work *on* them, so the work done on the gas is negative.

Note: Two sign conventions

Some books, especially in engineering, write the first law as $\Delta E_{\text{int}} = Q - W$, where W is the work done *by* the system. That W is the negative of ours, so the physics is identical; only the bookkeeping differs. We write the work done by a gas as $W_{\text{by}} = -W$ whenever we need it, so you will see both quantities, clearly labeled.

Checkpoint 37.2

(a) A gas absorbs $500\ \text{J}$ of heat while doing $200\ \text{J}$ of work on its surroundings. What is ΔE_{int}? (b) A gas is compressed, with $300\ \text{J}$ of work done on it, while it loses $100\ \text{J}$ of heat. What is ΔE_{int} now?

Answer

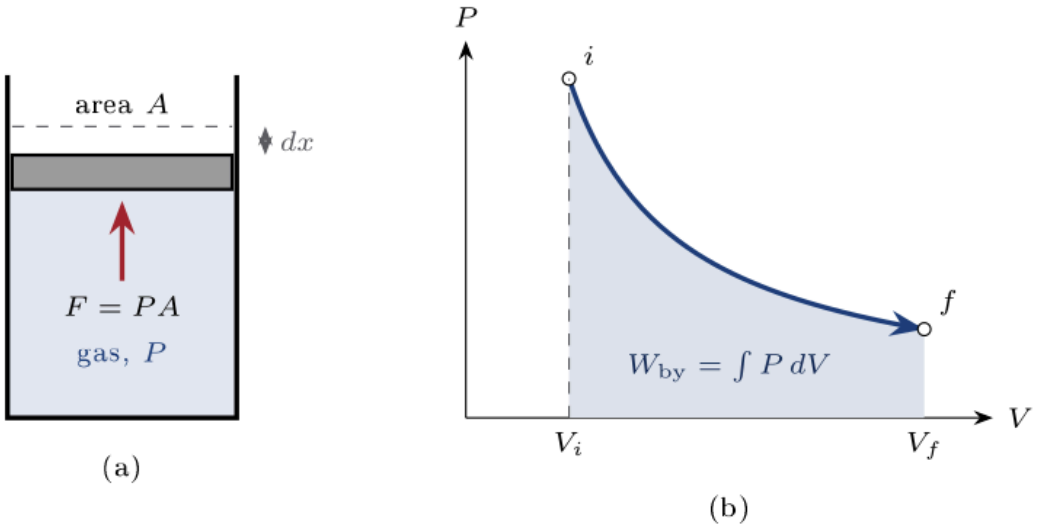


Figure 37.2. (a) A gas at pressure P pushes on a piston of area A with force PA. As the piston moves out by dx, the gas does work $PA\ dx = P\ dV$. (b) The total work done by the gas between states i and f is the area under its path on the P–V diagram.

•Figure description

Two panels. Panel a: a gas in a cylinder closed on top by a piston of area A. The gas pushes up on the piston with force P times A, and the piston rises by a small distance dx, so the gas volume grows by dV equal to A dx. Panel b: a pressure-versus-volume graph with a curved path from an initial state i at small volume and high pressure down to a final state f at larger volume and lower pressure. The region under the curve, between the curve and the volume axis from V sub i to V sub f, is shaded and labeled W sub by, the work done by the gas.

Work done by a gas. How much work does a gas do as it expands? Picture it in a cylinder, pushing on a piston of area A with a force $F = PA$ (Figure 37.2a). If the piston moves out a small distance dx, slowly enough that the pressure stays uniform, the gas does work $dW_{\text{by}} = F\ dx = PA\ dx$. But $A\ dx$ is the increase in the gas's volume, dV, so $dW_{\text{by}} = P\ dV$. For a finite change of volume,

Key Equation: Work done by a gas

$$W_{\text{by}} = -W = \int_{V_i}^{V_f} P\,dV. \tag{37.5}$$

This integral has a simple picture: it is the *area under the path* on a P–V diagram (Figure 37.2b). An expanding gas ($V_f > V_i$) does positive work; a compressed gas does negative work, meaning work is done on it.

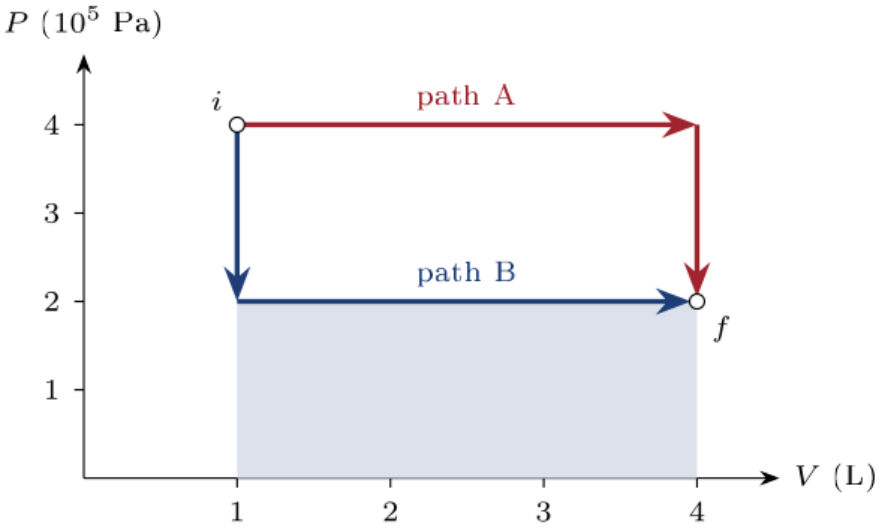


Figure 37.3. Two paths between the same initial and final states. The gas does more work along path A (area under its horizontal leg, $1200\ \text{J}$) than along path B (shaded area, $600\ \text{J}$), so it must also absorb more heat along path A. The change in internal energy is the same for both.

•Figure description

A pressure-versus-volume diagram with two states, i at volume 1.0 liter and pressure 4.0 times 10 to the 5 pascals, and f at volume 4.0 liters and pressure 2.0 times 10 to the 5 pascals. Path A, in red, runs horizontally right from i at the higher pressure to a corner at 4.0 liters, then straight down to f. Path B, in blue, runs straight down from i to a corner at 2.0 times 10 to the 5 pascals, then horizontally right to f. The area under path A, a tall rectangle, is larger than the area under path B, a short rectangle, which is shaded.

Because the work is an area under the path, it depends on the path, not just on where the gas starts and ends. Figure 37.3 shows two ways to take a gas from state i to state f. Along path A the gas expands first, at high pressure, and then cools at constant volume; along path B it cools first and then expands at low pressure. The gas does more work along path A. Its internal energy, though, is a state variable: ΔE_{int} depends only on the two end states, so it is the same along both paths. By the first law, $Q = \Delta E_{\text{int}} + W_{\text{by}}$, so the heat must also differ: along path A, the gas absorbs more heat to pay for the extra work. *Heat and work depend on the path; their combination $Q + W$ does not.* That is why it makes sense to talk about the internal energy a system *has* but not about the heat or work it has.

Example 37.7: Two paths, one change

A sample of monatomic ideal gas is taken from state i ($V_i = 1.0\ \text{L}$, $P_i = 4.0 \times 10^5\ \text{Pa}$) to state f ($V_f = 4.0\ \text{L}$, $P_f = 2.0 \times 10^5\ \text{Pa}$) along each path of Figure 37.3. For each path find the work done by the gas, the change in internal energy, and the heat absorbed.

Solution

Internal energy. For a monatomic ideal gas $E_{\text{int}} = \frac{3}{2}nRT = \frac{3}{2}PV$ (Chapter 36), so we never need n or T separately. With volumes in m^3,

$$\begin{aligned}\Delta E_{\text{int}} &= \tfrac{3}{2}(P_f V_f - P_i V_i)\\ &= \tfrac{3}{2}\left[(2.0\times10^5)(4.0\times10^{-3}) - (4.0\times10^5)(1.0\times10^{-3})\right]\\ &= \tfrac{3}{2}(800-400) = 600\,\text{J},\end{aligned} \tag{37.2.1}$$

the same for both paths.

Path A. Only the horizontal leg, at $4.0 \times 10^5\ \text{Pa}$, does work (the vertical leg has no volume change):

$$W_{\text{by,A}} = P\,\Delta V = (4.0\times10^5)(3.0\times10^{-3}) = \boxed{1200\,\text{J}}, \tag{37.2.2}$$

$$Q_A = \Delta E_{\text{int}} + W_{\text{by,A}} = 600 + 1200 = \boxed{1800\,\text{J}}. \tag{37.2.3}$$

Path B. Now the expansion happens at 2.0×10^5 Pa:

$$W_{\text{by,B}} = (2.0 \times 10^5)(3.0 \times 10^{-3}) = \boxed{600\ \text{J}}, \qquad Q_B = 600 + 600 = \boxed{1200\ \text{J}}. \tag{37.2.4}$$

Both paths raise the internal energy by 600 J. Path A takes in 600 J more heat and does exactly 600 J more work. Note that we used $W_{\text{by}} = Q - \Delta E_{\text{int}}$, the first law rearranged with the work done by the gas; with our sign convention, the work done *on* the gas is $W = -W_{\text{by}}$.

Two special cases follow directly from the first law. In a **cyclic process**, the system returns to its starting state, so $\Delta E_{\text{int}} = 0$ and $Q = W_{\text{by}}$: over a full cycle, the net heat absorbed equals the net work done. This is the principle of every heat engine (Chapter 38). And for an **isolated system**, which exchanges neither heat nor work with anything, $Q = W = 0$ and the internal energy cannot change at all.

Answers to Checkpoints

Checkpoint 37.2. (a) Here $Q = +500$ J (heat in), and the work done *on* the gas is $W = -200$ J, since the gas does work on its surroundings. So $\Delta E_{\text{int}} = Q + W = 500 - 200 = +300$ J. (b) Now $W = +300$ J and $Q = -100$ J (heat out), so $\Delta E_{\text{int}} = 300 - 100 = +200$ J. Keeping the signs straight is the whole problem: heat in and work done on the system both count as positive.

This page titled 37.2: The First Law of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.3: Thermodynamic Processes

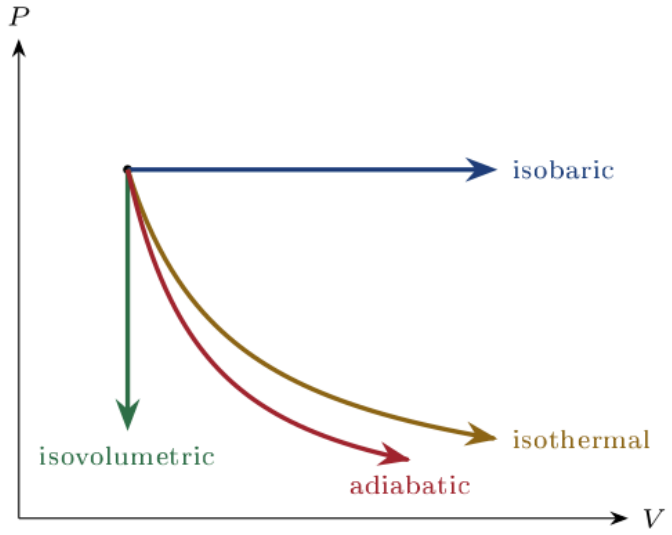


Figure 37.4. Four processes from a common starting state on a P–V diagram: *isobaric* (constant P), *isovolumetric* (constant V), *isothermal* (PV constant), and *adiabatic* ($Q = 0$, falling more steeply than the isotherm). The area under each path is the work done by the gas.

•Figure description

A pressure-versus-volume diagram with four paths leaving one common starting point at high pressure and low volume. The isobaric path runs horizontally to the right at constant pressure. The isovolumetric path runs straight down at constant volume. The isothermal path curves down and to the right along a hyperbola where the product P V stays constant. The adiabatic path leaves from the same point and curves down more steeply than the isothermal, falling below it. The area beneath each path is the work the gas does.

A change of state is a **thermodynamic process**, and four special cases are worth naming because in each one a different quantity is held fixed (Figure 37.4). For each, the first law tells us how the energy is shared between heat, work, and internal energy.

- **Isobaric**: constant pressure. The area under the horizontal path is a rectangle, so $W_{\text{by}} = P\ \Delta V$.
- **Isovolumetric** (isochoric): constant volume. No volume change means no work at all, so $\Delta E_{\text{int}} = Q$: all the heat goes into internal energy.
- **Isothermal**: constant temperature. The process follows an isotherm, $PV = \text{const}$. For an ideal gas, E_{int} depends only on T, so $\Delta E_{\text{int}} = 0$ and $Q = W_{\text{by}}$: every joule of heat absorbed leaves again as work.
- **Adiabatic**: no heat transfer, $Q = 0$. Then $\Delta E_{\text{int}} = W$; any change in internal energy comes entirely from work. A gas compressed adiabatically heats up, and a gas that expands adiabatically cools. Processes are nearly adiabatic when they are well insulated or simply too fast for much heat to flow, like the compression stroke of an engine or the expansion of air rising in the atmosphere.

Isothermal work. Along an isotherm the pressure changes as the gas expands, $P = nRT/V$, so we must integrate equation (37.5). With T constant,

$$W_{\text{by}} = \int_{V_i}^{V_f} \frac{nRT}{V}\, dV = nRT \int_{V_i}^{V_f} \frac{dV}{V}, \tag{37.3.1}$$

which gives

Key Equation: Work in an isothermal process

$$W_{\text{by}} = nRT \ln \frac{V_f}{V_i}. \tag{37.6}$$

The work is positive for an expansion ($V_f > V_i$, so the logarithm is positive) and negative for a compression.

Example 37.8: Boiling water: work, heat, and internal energy

Let $1.00\ \text{kg}$ of water at 100°C be boiled to steam at 100°C at standard atmospheric pressure ($1.01 \times 10^5\ \text{Pa}$). The volume rises from $1.00 \times 10^{-3}\ \text{m}^3$ (liquid) to $1.671\ \text{m}^3$ (steam). For water, $L_v = 2.26 \times 10^6\ \text{J/kg}$.

a. How much work does the system do during this process?
b. How much energy is transferred as heat?
c. What is the change in the system's internal energy?

Solution

(a) The process is isobaric, so the work done *by* the expanding steam is

$$W_{\text{by}} = P\,\Delta V = (1.01 \times 10^5)(1.671 - 0.001) = (1.01 \times 10^5)(1.670) \approx \boxed{169\,\text{kJ}}. \tag{37.3.2}$$

(b) The heat is the latent heat of vaporization:

$$Q = mL_v = (1.00)(2.26 \times 10^6) = 2.26 \times 10^6\,\text{J} = \boxed{2260\,\text{kJ}}. \tag{37.3.3}$$

(c) By the first law, with the work done *on* the system being $W = -W_{\text{by}} = -169\ \text{kJ}$,

$$\Delta E_{\text{int}} = Q + W = 2260 - 169 \approx \boxed{2090\,\text{kJ}}. \tag{37.3.4}$$

Of the $2260\ \text{kJ}$ poured in, about $169\ \text{kJ}$ went into pushing back the atmosphere as the steam expanded, and the remaining $2090\ \text{kJ}$ went into the internal energy of the vapor: mostly the energy of pulling the water molecules apart.

Checkpoint 37.3

An ideal gas starts in a given state and expands to twice its volume, either (a) at constant pressure or (b) at constant temperature. Sketch both paths on a P–V diagram. In which process does the gas do more work, and which requires more heat?

Answer

Molar Specific Heats of an Ideal Gas

For a gas it is more natural to count moles than kilograms, so we define the **molar specific heat** C by $Q = nC\ \Delta T$. Unlike a solid or liquid, though, a gas has no single specific heat, because the heat needed for a given ΔT depends on the process. Two cases matter most.

At constant volume the gas does no work, so all the heat goes into internal energy: $Q = \Delta E_{\text{int}}$. From the equipartition theorem, $E_{\text{int}} = \frac{f}{2}nRT$ (Chapter 36), so $\Delta E_{\text{int}} = \frac{f}{2}nR\ \Delta T$, and the molar specific heat at constant volume is

$$C_V = \frac{f}{2}R = \begin{cases} \frac{3}{2}R = 12.5\,\text{J/(mol}\cdot\text{K)} & \text{monatomic,} \\ \frac{5}{2}R = 20.8\,\text{J/(mol}\cdot\text{K)} & \text{diatomic.} \end{cases} \tag{37.7}$$

Because the internal energy of an ideal gas depends only on its temperature, this gives a result we will use constantly:

$$\Delta E_{\text{int}} = nC_V\,\Delta T \qquad \text{for } \textit{any} \text{ process of an ideal gas.} \tag{37.8}$$

At constant pressure the gas expands as it warms, so the heat must also pay for the work it does. The first law gives $Q = \Delta E_{\text{int}} + W_{\text{by}} = nC_V\ \Delta T + P\ \Delta V$. At constant pressure the ideal gas law gives $P\ \Delta V = nR\ \Delta T$, so $Q = n(C_V + R)\ \Delta T$, and the molar specific heat at constant pressure is

$$C_P = C_V + R. \tag{37.9}$$

Their ratio, $\gamma = C_P/C_V$, is the **adiabatic index**: $\frac{5}{3} \approx 1.67$ for a monatomic gas and $\frac{7}{5} = 1.40$ for a diatomic one.

Table 37.3. Measured molar specific heats of gases near room temperature, compared with the equipartition predictions: $C_V = \frac{3}{2}R = 12.5$ and $\gamma = 1.67$ (monatomic), $C_V = \frac{5}{2}R = 20.8$ and $\gamma = 1.40$ (diatomic). Units of C_V and C_P are $\mathrm{J/(mol \cdot K)}$.

Gas	Type	C_V	C_P	$\gamma = C_P/C_V$
Helium (He)	monatomic	12.5	20.8	1.67
Argon (Ar)	monatomic	12.5	20.8	1.67
Nitrogen (N_2)	diatomic	20.8	29.1	1.40
Oxygen (O_2)	diatomic	21.1	29.4	1.40
Carbon dioxide (CO_2)	triatomic	28.5	37.0	1.30

Table 37.3 shows how well this works. The monatomic gases match the prediction exactly, and nitrogen and oxygen agree to within about 1%, confirming that their molecules rotate but, at room temperature, do not vibrate (the frozen-out vibration of Chapter 36). Every gas obeys $C_P - C_V = R = 8.3\ \mathrm{J/(mol \cdot K)}$.

Adiabatic processes. When $Q = 0$, the first law says $dE_{\text{int}} = -P\ dV$, or $nC_V\ dT = -P\ dV$. Differentiating $PV = nRT$ gives $P\ dV + V\ dP = nR\ dT$. Eliminating dT between these two equations and using $R = C_P - C_V$ leads, after a few lines of algebra, to $dP/P = -\gamma\ dV/V$, which integrates to

Key Equation: Adiabatic process of an ideal gas

$$PV^{\gamma} = \text{constant}, \qquad TV^{\gamma-1} = \text{constant}. \tag{37.10}$$

(The second form follows from the first with $P = nRT/V$.) Since $\gamma > 1$, an adiabat falls more steeply than an isotherm on a P–V diagram, as Figure 37.4 shows: as the gas expands, its pressure drops both because the volume grows and because the gas cools. The work done by the gas in an adiabatic process comes entirely out of its internal energy, $W_{\text{by}} = -\Delta E_{\text{int}} = nC_V(T_i - T_f)$.

Example 37.9: The diesel engine

A diesel engine has no spark plugs. Air is drawn into the cylinder at $20^{\circ}\mathrm{C}$ and $1.01 \times 10^5\ \mathrm{Pa}$ and compressed so rapidly that the process is essentially adiabatic, to $1/18$ of its original volume. Fuel is then sprayed in. Find the temperature and pressure of the compressed air. Treat air as a diatomic ideal gas, $\gamma = 1.40$.

Solution

From $TV^{\gamma-1} = \text{const}$,

$$T_f = T_i \left(\frac{V_i}{V_f}\right)^{\gamma-1} = (293\,\mathrm{K})(18)^{0.40} = (293)(3.18) = \boxed{932\,\mathrm{K}} \approx 659^{\circ}\mathrm{C}. \tag{37.3.5}$$

From $PV^{\gamma} = \text{const}$,

$$P_f = P_i \left(\frac{V_i}{V_f}\right)^{\gamma} = (1.01 \times 10^5)(18)^{1.40} = (1.01 \times 10^5)(57.2) = \boxed{5.8 \times 10^6\,\mathrm{Pa}}, \tag{37.3.6}$$

about 57 atmospheres. Diesel fuel ignites spontaneously at around $250^{\circ}\mathrm{C}$, so the hot compressed air sets it burning the moment it is injected: compression alone does the job of the spark. (The pressure rises by a factor of 57 while the volume falls by only 18, because the gas also heats up.)

Checkpoint 37.4

Pumping up a bicycle tire quickly makes the barrel of the pump warm to the touch, even though no heat is supplied. Explain why, using the first law. Would the pump warm as much if you pumped very slowly?

Answer

Answers to Checkpoints

Checkpoint 37.3. The work done by the gas is the area under its path on the P–V diagram. At constant pressure the path is a horizontal line from V to $2V$; at constant temperature the pressure falls along an isotherm ($P \propto 1/V$) and ends at half its starting value. The isotherm lies below the horizontal line everywhere after the start, so the constant-pressure expansion encloses more area: the gas does more work, P_0V_0 versus $P_0V_0 \ln 2 \approx 0.69\ P_0V_0$. It also needs more heat. In the isothermal case $\Delta E_{\text{int}} = 0$, so Q just equals the work. At constant pressure the temperature doubles ($PV = nRT$), so heat must supply both the larger work *and* an increase in internal energy.

Checkpoint 37.4. Each quick stroke compresses the air in the pump faster than heat can leak out, so the compression is nearly adiabatic: $Q \approx 0$ and $\Delta E_{\text{int}} = W > 0$. The work you do on the air raises its internal energy, and so its temperature, and the hot air warms the barrel. (Friction of the piston adds a little.) Pumped very slowly, the air would stay close to room temperature, because heat would have time to flow out through the barrel during each stroke; the compression would be nearly isothermal and the pump would barely warm.

This page titled 37.3: Thermodynamic Processes is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.4: Heat Transfer Mechanisms

Heat moves from hot to cold by three distinct mechanisms: **conduction**, through direct contact; **convection**, carried by a moving fluid; and **radiation**, carried by electromagnetic waves. Most real situations involve all three, but each has its own rules.

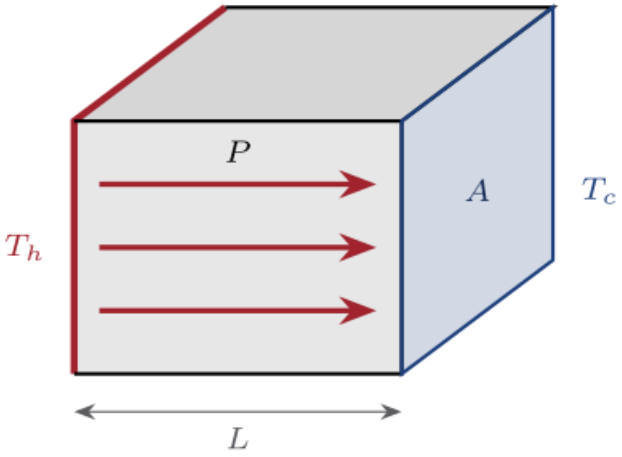


Figure 37.5. Conduction through a slab. Heat flows from the hot face at T_h to the cold face at T_c at a rate $P = kA(T_h - T_c)/L$, proportional to the area and the temperature difference and inversely proportional to the thickness.

•Figure description

A slab of material drawn in perspective, its thickness L running left to right. The heat-flow faces are the left and right faces, each of area A. The right face, held at the lower temperature T sub c, is visible, shaded blue and labeled A. The left face, held at the higher temperature T sub h, faces away from the viewer, so only its two visible edges are drawn, in red. Three arrows labeled P, the rate of heat flow, cross the front of the slab from the hot face to the cold face.

Conduction. In **conduction**, energy passes through a material as its faster-jiggling molecules or atoms bump into slower neighbors, with no bulk motion of the material at all. A metal spoon left in hot soup grows warm at the handle by conduction. For a slab of area A and thickness L whose faces are held at temperatures T_h and T_c (Figure 37.5), experiment shows that the rate of heat flow, the energy per unit time, is

Key Equation: Heat conduction

$$P_{\text{cond}} = \frac{Q}{t} = kA\,\frac{T_h - T_c}{L}. \tag{37.11}$$

The constant k is the **thermal conductivity** of the material (Table 37.4), in $\mathrm{W/(m \cdot K)}$. It ranges over five orders of magnitude. Metals conduct heat best, because their free electrons (the same ones that carry electric current, Chapter 42) carry energy quickly through the solid. Gases conduct worst, because their molecules are far apart. Nearly every good insulator, from wool and down feathers to foam and fiberglass, works the same way: it traps air in tiny pockets where it cannot circulate, and lets air's low conductivity do the rest.

Table 37.4. Thermal conductivities near room temperature.

Material	k ($\mathrm{W/(m \cdot K)}$)	Material	k ($\mathrm{W/(m \cdot K)}$)
Silver	427	Water	0.60
Copper	397	Wood (pine)	0.12
Aluminum	238	Fiberglass insulation	0.040
Steel	50	Polystyrene foam	0.033
Glass	0.84	Air	0.026

Material	k ($\mathrm{W/(m\cdot K)}$)	Material	k ($\mathrm{W/(m\cdot K)}$)
Concrete	0.80		

When heat must cross several layers in a row, such as the wood, insulation, and drywall of a house wall, the same heat flow passes through each layer in turn. The temperature drops across the layers add up to the total, and the result is

$$P_{\text{cond}} = \frac{A(T_h - T_c)}{\sum_i L_i/k_i}. \tag{37.12}$$

Each term L_i/k_i is the layer's **thermal resistance** per unit area (sold in the building trade as an "R-value"), and resistances in series simply add, just as electrical resistances do in Chapter 21.

Example 37.10: A picnic cooler

A polystyrene-foam cooler has a total wall area of $0.80\ \mathrm{m^2}$ and walls $2.0\ \mathrm{cm}$ thick. It holds ice at $0^\circ\mathrm{C}$ on a $30^\circ\mathrm{C}$ day. (a) At what rate does heat leak in through the walls? (b) How long will $5.0\ \mathrm{kg}$ of ice last?

Solution

(a) Heat flow. With $k = 0.033\ \mathrm{W/(m\cdot K)}$ from Table 37.4,

$$P = kA\,\frac{T_h - T_c}{L} = (0.033)(0.80)\frac{30}{0.020} = \boxed{40\,\mathrm{W}}. \tag{37.4.1}$$

(b) Melting time. Melting the ice takes $Q = mL_f = (5.0)(3.33\times10^5) = 1.67\times10^6\ \mathrm{J}$, so

$$t = \frac{Q}{P} = \frac{1.67\times10^6\,\mathrm{J}}{40\,\mathrm{W}} = 4.2\times10^4\,\mathrm{s} \approx \boxed{12\,\mathrm{h}}. \tag{37.4.2}$$

The ice lasts about half a day. In practice it lasts a little less, because heat also leaks in each time the lid is opened and the warm air inside is replaced.

Example 37.11: Insulating a wall

A $10.0\ \mathrm{m^2}$ section of wall is made of $2.0\ \mathrm{cm}$ of pine boards. The inside is at $20^\circ\mathrm{C}$ and the outside at $-5^\circ\mathrm{C}$. (a) Find the rate of heat loss. (b) Find the rate of heat loss after a $9.0\ \mathrm{cm}$ layer of fiberglass insulation is added.

Solution

(a) Wood alone.

$$P = kA\,\frac{\Delta T}{L} = (0.12)(10.0)\frac{25}{0.020} = \boxed{1500\,\mathrm{W}}. \tag{37.4.3}$$

(b) Wood plus fiberglass. The thermal resistances add:

$$\sum \frac{L_i}{k_i} = \frac{0.020}{0.12} + \frac{0.090}{0.040} = 0.17 + 2.25 = 2.42\,\mathrm{m^2\cdot K/W}, \tag{37.4.4}$$

$$P = \frac{A\,\Delta T}{\sum L_i/k_i} = \frac{(10.0)(25)}{2.42} = \boxed{103\,\mathrm{W}}. \tag{37.4.5}$$

The insulation cuts the heat loss by a factor of about fifteen. Nearly all of the thermal resistance is in the fiberglass, the layer with the largest L/k. (These figures ignore the thin layers of still air clinging to each surface, which add a little more resistance in real walls.)

Convection. In **convection**, heat is carried by the bulk motion of a fluid. Air warmed by a radiator expands (Chapter 36), becomes less dense than the cooler air around it, and rises; cooler air flows in to take its place, and a circulating current carries energy around the room. This is *natural* convection, driven by buoyancy (Chapter 14). *Forced* convection uses a fan or pump instead, as

in a car's radiator or a convection oven, and it carries heat away much faster, which is why a breeze feels cooler than still air at the same temperature. Convection on a planetary scale moves heat from the tropics toward the poles through winds and ocean currents, and it drives the daily sea breeze: the land heats up faster than the sea during the day, the air above it rises, and cool air flows in off the water.

Radiation. In **radiation**, energy is carried by electromagnetic waves (Chapter 27), so no medium is needed at all. The Sun warms the Earth across $150\ \text{million}$ kilometers of empty space by radiation. Every object emits thermal radiation, at a rate given by the **Stefan–Boltzmann law**:

Key Equation: Stefan–Boltzmann law

$$P_{\text{rad}} = \sigma \varepsilon A T^4. \tag{37.13}$$

Here A is the surface area, T the Kelvin temperature of the surface, and $\sigma = 5.670 \times 10^{-8}\ \text{W}/(\text{m}^2 \cdot \text{K}^4)$ the **Stefan–Boltzmann constant**. The **emissivity** ε, between 0 and 1, describes the surface: a perfect absorber, an ideal *blackbody*, has $\varepsilon = 1$, and a shiny reflective surface has ε near 0. A good absorber is always an equally good emitter. The steep T^4 means radiation dominates at high temperatures: doubling the Kelvin temperature multiplies the emitted power by sixteen. An object also absorbs radiation from its surroundings, at temperature T_{env}, so its *net* rate of radiative loss is

$$P_{\text{net}} = \sigma \varepsilon A \left(T^4 - T_{\text{env}}^4\right). \tag{37.14}$$

At room temperature the radiation is almost all infrared, invisible to us but plain to a thermal camera; at about $800\ \text{K}$ an object begins to glow a dull red. Explaining the colors of this glow was the problem that launched quantum mechanics (Chapter 39).

Example 37.12: Radiating like a light bulb

A person with a skin area of $1.5\ \text{m}^2$ and skin temperature 34°C sits unclothed in a room whose walls are at 22°C. Skin, whatever its color, is nearly a blackbody in the infrared, $\varepsilon = 0.97$. Find the net rate of heat loss by radiation.

Solution

In kelvins, $T = 307\ \text{K}$ and $T_{\text{env}} = 295\ \text{K}$. Then

$$T^4 - T_{\text{env}}^4 = (307)^4 - (295)^4 = 8.88 \times 10^9 - 7.57 \times 10^9 = 1.31 \times 10^9\,\text{K}^4, \tag{37.4.6}$$

$$P_{\text{net}} = (5.67 \times 10^{-8})(0.97)(1.5)(1.31 \times 10^9) \approx \boxed{110\,\text{W}}. \tag{37.4.7}$$

That is about as much as the body's entire resting metabolic output of roughly $100\ \text{W}$, which is why an unclothed person in a 22°C room soon feels chilly even though the air is not cold. Clothing helps less by blocking radiation than by trapping a layer of warm air next to the skin. Note that we had to use Kelvin temperatures and subtract their fourth powers; subtracting first, $(307 - 295)^4$, would be badly wrong.

A good vacuum bottle defeats all three mechanisms at once: a vacuum gap between its walls stops conduction and convection (no matter to carry the heat), and a shiny reflective lining, with ε close to zero, turns back radiation. That is why such a bottle keeps coffee hot and lemonade cold for hours.

Checkpoint 37.5

On a chilly night, a tile floor feels much colder to your bare feet than a wool rug in the same room, even though both are at the same temperature. Which heat-transfer mechanism explains the difference, and in which direction is the heat flowing?

Answer

Answers to Checkpoints

Checkpoint 37.5. Conduction. Both surfaces are at the same temperature, below that of your skin, so in both cases heat flows

from your feet *into* the floor. Tile has a much higher thermal conductivity than wool (which traps insulating air), so it draws heat out of your feet far faster. Your skin senses the rate of heat loss, not the floor's temperature, so the tile feels colder.

This page titled 37.4: Heat Transfer Mechanisms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.S: Heat and the First Law of Thermodynamics (Summary)

Summary

- **Heat** Q is energy transferred because of a temperature difference, always flowing hot $\rightarrow$ cold. It is distinct from temperature (average molecular kinetic energy) and from internal energy (the total energy a system holds). $1\ \text{cal} = 4.186\ \text{J}$; a food Calorie is $4186\ \text{J}$.
- To change temperature, $Q = mc\ \Delta T$; water's large specific heat moderates coastal climates. To change phase, $Q = mL$, with the temperature constant during the change. **Calorimetry**: $\sum Q = 0$ for an isolated system.
- **First law**: $\Delta E_{\text{int}} = Q + W$, with heat into the system and work done *on* the system positive. A gas does work $W_{\text{by}} = -W = \int P\ dV$, the area under its path on a P–V diagram. Q and W depend on the path; ΔE_{int} does not. In a cycle, $\Delta E_{\text{int}} = 0$.
- **Processes**: isobaric ($W_{\text{by}} = P\ \Delta V$), isovolumetric ($W = 0$, $\Delta E_{\text{int}} = Q$), isothermal ($\Delta E_{\text{int}} = 0$, $W_{\text{by}} = nRT\ln(V_f/V_i)$), adiabatic ($Q = 0$, PV^{γ} and $TV^{\gamma-1}$ constant).
- **Molar specific heats** of an ideal gas: $C_V = \frac{f}{2}R$, $C_P = C_V + R$, $\gamma = C_P/C_V$ ($\frac{5}{3}$ monatomic, $\frac{7}{5}$ diatomic). For any process, $\Delta E_{\text{int}} = nC_V\ \Delta T$.
- **Heat transfer**: *conduction*, $P = kA\ \Delta T/L$, with layer resistances L/k adding in series; *convection*, by fluid motion; and *radiation*, $P = \sigma\varepsilon AT^4$, with net loss $\sigma\varepsilon A(T^4 - T_{\text{env}}^4)$.

Key Equations

Heat and temperature change

$$Q = mc\,\Delta T \tag{37.S.1}$$

Heat of a phase change

$$Q = mL \qquad (L_f \text{ fusion},\ L_v \text{ vaporization},\ L_s \text{ sublimation}) \tag{37.S.2}$$

Calorimetry

$$\sum Q = 0 \tag{37.S.3}$$

First law of thermodynamics

$$\Delta E_{\text{int}} = Q + W \tag{37.S.4}$$

Work done by a gas

$$W_{\text{by}} = -W = \int_{V_i}^{V_f} P\,dV \qquad (\text{isobaric: } W_{\text{by}} = P\,\Delta V) \tag{37.S.5}$$

Work in an isothermal process

$$W_{\text{by}} = nRT\ln\frac{V_f}{V_i} \tag{37.S.6}$$

Molar specific heats of an ideal gas

$$C_V = \frac{f}{2}R, \qquad C_P = C_V + R, \qquad \gamma = \frac{C_P}{C_V}, \qquad \Delta E_{\text{int}} = nC_V\,\Delta T \tag{37.S.7}$$

Adiabatic process

$$PV^{\gamma} = \text{const}, \qquad TV^{\gamma-1} = \text{const} \tag{37.S.8}$$

Heat conduction

$$P_{\text{cond}} = kA\,\frac{T_h - T_c}{L} = \frac{A(T_h - T_c)}{\sum L_i/k_i} \tag{37.S.9}$$

Stefan–Boltzmann law

$$P_{\text{rad}} = \sigma\varepsilon A T^4, \qquad P_{\text{net}} = \sigma\varepsilon A\left(T^4 - T_{\text{env}}^4\right) \tag{37.S.10}$$

This page titled 37.S: Heat and the First Law of Thermodynamics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

37.E: Heat and the First Law of Thermodynamics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $c_{\text{water}} = 4186\ \text{J/(kg·K)}$, $c_{\text{ice}} = 2090\ \text{J/(kg·K)}$, $L_f = 3.33 \times 10^5\ \text{J/kg}$, $L_v = 2.26 \times 10^6\ \text{J/kg}$, $R = 8.314\ \text{J/(mol·K)}$, $\sigma = 5.67 \times 10^{-8}\ \text{W/(m}^2\text{·K}^4\text{)}$, Tables 37.1–37.4, and other constants as given.

Section 37.1 — Heat: Temperature Changes and Phase Changes

37.1 ★ How much heat is required to raise the temperature of $2.0\ \text{kg}$ of water from 20°C to 80°C?

37.2 ★ *[History of science.]* Joule is said to have measured the water temperature at the top and bottom of a waterfall on his honeymoon. If all of the gravitational potential energy of water falling $50\ \text{m}$ became internal energy of the water, how much warmer would it be at the bottom? Why was this a hard experiment?

37.3 ★★ A $0.50\ \text{kg}$ block of aluminum ($c = 900\ \text{J/(kg·K)}$) absorbs $2.7 \times 10^4\ \text{J}$ of heat. By how much does its temperature rise?

37.4 ★★ How much heat must be removed from $0.30\ \text{kg}$ of water at 20°C to freeze it completely into ice at 0°C?

37.5 ★★ *[Real-world application: biology.]* Sweating cools the body by evaporation. Estimate the heat removed when $0.20\ \text{kg}$ of sweat (treat as water) evaporates from the skin at body temperature.

37.6 ★★ How much heat is required to take $0.50\ \text{kg}$ of water from 20°C all the way to steam at 100°C?

37.7 ★★ A $0.30\ \text{kg}$ block of ice at 0°C is added to a drink. How much heat must be supplied to melt the ice completely and then warm the meltwater to 10°C?

Section 37.1.3 — Calorimetry

37.8 ★★ A $0.25\ \text{kg}$ piece of copper ($c = 387\ \text{J/(kg·K)}$) at 150°C is dropped into $0.40\ \text{kg}$ of water at 20°C in an insulated cup of negligible heat capacity. Find the equilibrium temperature.

37.9 ★★★ A $0.10\ \text{kg}$ ice cube at 0°C is dropped into $0.50\ \text{kg}$ of water at 30°C in an insulated container.

a. Show that all the ice melts.
b. Find the final temperature of the mixture.

37.10 ★★ A $200\ \text{g}$ block of metal at 95°C is dropped into $300\ \text{g}$ of water at 20°C in an insulated cup, reaching a final temperature of 26°C. Neglecting the cup, find the specific heat of the metal.

Sections 37.2–37.3 — The First Law and Thermodynamic Processes

37.11 ★ *[Conceptual.]* For each process, state what is zero (Q, W, or ΔE_{int}): (a) an adiabatic compression; (b) an isovolumetric heating; (c) the isothermal process of an ideal gas (recall E_{int} depends only on T).

37.12 ★★ A gas is compressed isobarically at $1.5 \times 10^5\ \text{Pa}$ from $0.080\ \text{m}^3$ to $0.050\ \text{m}^3$, and in the process it releases $8.0 \times 10^3\ \text{J}$ of heat.

a. Find the work done on the gas.
b. Find the change in internal energy.

37.13 ★★ An ideal gas is compressed, with $400\ \text{J}$ of work done *on* it while it releases $250\ \text{J}$ of heat to its surroundings. Find the change in its internal energy.

37.14 ★★ $2.0\ \text{mol}$ of an ideal gas expands isothermally at $350\ \text{K}$ to twice its original volume.

a. Find the work done by the gas.
b. How much heat flows, and in which direction?

37.15 ★★ A monatomic ideal gas is taken from state i ($1.0\ \text{L}$, $2.0 \times 10^5\ \text{Pa}$) to state f ($3.0\ \text{L}$, $2.0 \times 10^5\ \text{Pa}$) by two different paths. Path A is a single isobaric expansion. Path B first raises the pressure at constant volume to $4.0 \times 10^5\ \text{Pa}$, then expands isobarically to $3.0\ \text{L}$, then lowers the pressure at constant volume to $2.0 \times 10^5\ \text{Pa}$. (a) Sketch both paths on a P–V diagram. (b) For each path, find the work done by the gas, the change in internal energy, and the heat absorbed. (c) Which quantities are the same for both paths, and why?

37.16 ★★ $2.0\ \text{mol}$ of nitrogen gas is heated from $300\ \text{K}$ to $350\ \text{K}$. Using Table 37.3, find the heat required (a) at constant volume and (b) at constant pressure. (c) Show that the difference equals the work the gas does in case (b).

37.17 ★★ *[Real-world application: the bicycle pump.]* Air at 20°C and $1.0 \times 10^5\ \text{Pa}$ is compressed quickly, and so adiabatically, to one third of its original volume. Treat air as a diatomic ideal gas. Find (a) its final temperature and (b) its final pressure. (c) Compare with the final pressure if the same compression were done slowly enough to be isothermal.

Section 37.4 — Heat Transfer Mechanisms & the Coastal Climate

37.18 ★ *[Conceptual.]* Name the dominant heat-transfer mechanism in each: (a) a hand warmed near (not touching) a campfire; (b) a metal railing chilling your hand on a cold day; (c) a hot-air balloon rising.

37.19 ★★ A copper rod $0.50\ \text{m}$ long with a cross-sectional area of $1.0\ \text{cm}^2$ is insulated along its sides. One end is held in boiling water at 100°C and the other in an ice–water mixture at 0°C. (a) At what rate does heat flow along the rod? (b) How many grams of ice melt per hour?

37.20 ★★ A window is a single pane of glass $4.0\ \text{mm}$ thick. A double-glazed replacement has two $3.0\ \text{mm}$ panes separated by a $12\ \text{mm}$ gap of still air. (a) Find the thermal resistance per unit area, $\sum L/k$, of each. (b) By what factor does the double pane reduce the heat flow, according to these numbers? (c) Measured improvements are much smaller, closer to a factor of three. Suggest two reasons.

37.21 ★★ *[Real-world application: astronomy.]* The Sun is very nearly a blackbody ($\varepsilon = 1$) with a radius of $6.96 \times 10^8\ \text{m}$ and a surface temperature of $5772\ \text{K}$. (a) Find its total power output. (b) Find the intensity of sunlight at Earth's distance, $1.50 \times 10^{11}\ \text{m}$, and compare with the measured value of about $1361\ \text{W/m}^2$.

37.22 ★★★ *[Real-world application: the coastal climate.]* Reread the chapter opener. A coastal town sits beside the ocean; an inland town sits beside dry soil ($c_{\text{soil}} \approx 800\ \text{J/(kg}\cdot\text{K)}$).

a. The same midday sunlight delivers the same heat Q to equal masses of seawater and of dry soil. Using $\Delta T = Q/(mc)$, find the ratio of the soil's temperature rise to the water's.
b. Explain in one or two sentences how this ratio accounts for the mild coastal climate and the extreme inland one.
c. At night both surfaces radiate heat away. Will the coast or the inland town cool more, and why?

Challenge Problems

37.23 ★★★ *[Real-world application: energy — a solar water heater.]* A rooftop collector of area $4.0\ \text{m}^2$ receives sunlight at an intensity of $800\ \text{W/m}^2$ and delivers 62% of that energy to the water in a $150\ \text{L}$ tank. Take $c_{\text{water}} = 4186\ \text{J/(kg}\cdot\text{K)}$ and $\rho_{\text{water}} = 1000\ \text{kg/m}^3$.

a. How much heat is required to raise the tank from 15°C to 55°C?
b. What useful power does the collector deliver?
c. How long does the heating take, in hours?
d. Water's volume expansion coefficient near room temperature is $\beta \approx 2.1 \times 10^{-4}\ \text{K}^{-1}$. By how many liters does the tank's water expand over that temperature rise? Explain why a sealed solar hot-water system must include an expansion tank, and estimate what would happen to the pressure without one.

37.24 ★★★ *[Real-world application: engineering — an air-compressor cycle.]* $2.0\ \text{mol}$ of a monatomic ideal gas is taken around the closed cycle $A \to B \to C \to A$:

- $A \to B$: isobaric expansion at $P = 1.0 \times 10^5\ \text{Pa}$ from $V_A = 0.050\ \text{m}^3$ to $V_B = 0.100\ \text{m}^3$;
- $B \to C$: isovolumetric cooling back to the starting temperature;
- $C \to A$: isothermal compression back to V_A.

a. Find the temperature at each of the three states.
b. For each leg, find the work done *by* the gas, the change in internal energy, and the heat transferred. (Recall $E_{\text{int}} = \frac{3}{2}nRT$ for a monatomic gas, and $W_{\text{by}} = nRT\ln(V_f/V_i)$ for an isothermal step.)
c. Verify that $\Delta E_{\text{int}} = 0$ around the complete cycle, and that the net work done by the gas equals the net heat absorbed.
d. Sketch the cycle on a P–V diagram, state whether it runs clockwise or counterclockwise, and compute the ratio of the net work to the heat absorbed on the $A \to B$ leg. What does that ratio represent?

This page titled 37.E: Heat and the First Law of Thermodynamics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38: ENTROPY AND THE SECOND LAW OF THERMODYNAMICS

Drop a ball and it bounces, a little lower each time, until it lies still on the floor. You have seen this ten thousand times and never thought it strange. But now ask the reverse: have you ever seen a ball lying still on the floor suddenly gather itself up and begin to bounce, climbing higher with each rebound? Never, and yet nothing in the laws of mechanics forbids it. Energy would be perfectly conserved if the floor's thermal energy flowed back into the ball and launched it upward. The first law of thermodynamics, which we built in Chapter 37, has no objection. Something *else* decides that the movie of the bouncing ball may run only forward, never backward.

That something is **entropy**, and the rule it obeys (the **second law of thermodynamics**) is one of the deepest statements in all of physics. It is the law that gives time its direction, that explains why heat flows from hot to cold and never the reverse, and that sets an absolute ceiling on how efficient any engine can ever be. By the end of this chapter you will understand entropy from two complementary angles: the macroscopic one, where entropy is heat divided by temperature, and the far more revealing statistical one, where entropy counts the number of microscopic arrangements a system can have. You will see that the second law is, at bottom, not a law of force at all but a law of overwhelming probability, and you will be able to use it to find the best efficiency any heat engine can achieve.

Learning Objectives

- Calculate the entropy change of a reversible process from $dS = dQ_r/T$.
- Interpret entropy statistically through microstates, macrostates, and multiplicity, and connect the two views with $S = k_B \ln W$.
- Distinguish reversible from irreversible processes and compute entropy changes for real (irreversible) processes by routing through a reversible path.
- State the second law of thermodynamics in both its forms.
- Analyze heat engines and compute their efficiency, including the maximum (Carnot) efficiency set by the reservoir temperatures.
- Read the net work, the heat absorbed, and the efficiency of a cycle from its P–V diagram.
- Apply the same ideas in reverse to refrigerators.

We will return to the bouncing ball, and the arrow of time, once we have the statistical view, and again in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

38.1: Entropy
38.2: The Second Law of Thermodynamics
38.3: Heat Engines
38.S: Entropy and the Second Law of Thermodynamics (Summary)
38.E: Entropy and the Second Law of Thermodynamics (Problems)

This page titled 38: Entropy and the Second Law of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38.1: Entropy

To make the idea precise we start macroscopically. The **entropy** S of a system is a state variable: like pressure or internal energy, it depends only on the current state, not on how the system got there. Its *change* during a reversible process is defined by

Key Equation: Entropy change (reversible)

$$dS = \frac{dQ_r}{T} \qquad \Longrightarrow \qquad \Delta S = \int \frac{dQ_r}{T}, \tag{38.1}$$

where dQ_r is the heat added in a *reversible* step and T is the (kelvin) temperature at which it is added. The units of entropy are joules per kelvin ($\mathrm{J/K}$). Adding heat to a system raises its entropy; removing heat lowers it. And because the temperature sits in the denominator, the *same* amount of heat changes the entropy more when added at a low temperature than at a high one: a point that will turn out to be the secret behind the second law.

That entropy is a state function is the key that unlocks everything. If a process carries a system from one state to another and back again, the total entropy change must be zero, since it returns to the same state. We will exploit this repeatedly.

A Statistical View of Entropy

The macroscopic definition tells us how to compute entropy but not what it *is*. The answer comes from counting. Distinguish two ways of describing a system. A **macrostate** is the large-scale description: the pressure, volume, and temperature you could actually measure. A **microstate** is the complete microscopic description: the exact position and motion of every single particle. Many different microstates look identical at the macroscopic level, and the number of microstates that correspond to a given macrostate is its **multiplicity** W.

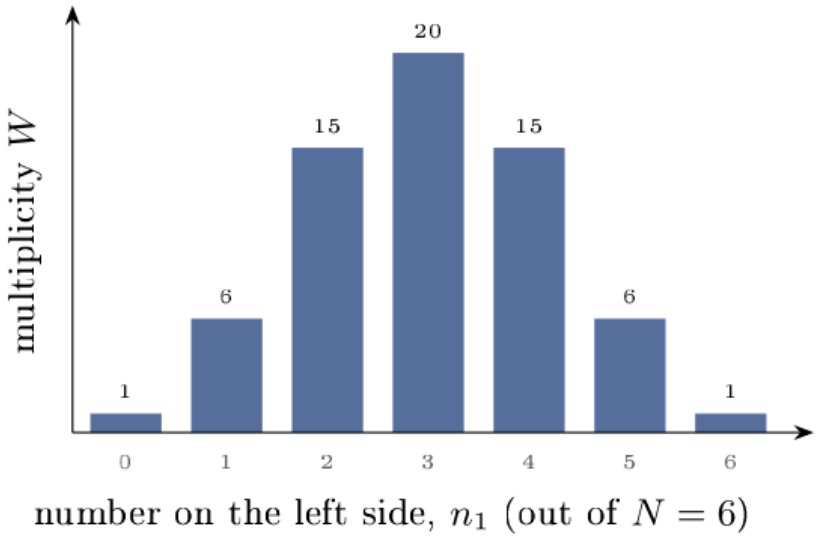


Figure 38.1. Multiplicity for $N = 6$ particles in a divided box. The evenly split macrostate ($n_1 = 3$) has the most microstates ($W = 20$); the lopsided ones have far fewer, and "all on one side" has only $W = 1$. For real $N \sim 10^{23}$ the central peak becomes overwhelmingly sharp.

•Figure description

A bar chart of multiplicity W against the number of particles on the left-hand side of a divided box, for six particles in total. The seven bars, from zero on the left through six, have heights 1, 6, 15, 20, 15, 6 and 1. The chart is symmetric and peaks sharply in the middle at the evenly split macrostate of three and three, which has twenty microstates, while each all-on-one-side macrostate has only one.

For a simple model (N particles free to be on either side of a divided box) the multiplicity of the macrostate with n_1 on the left and n_2 on the right is the number of ways to choose which particles go where,

$$W = \frac{N!}{n_1!\, n_2!}. \tag{38.2}$$

Try it with a handful of particles and a pattern emerges (Figure 38.1): the "evenly split" macrostate, $n_1 = n_2$, has by far the largest multiplicity, while the "all on one side" macrostates have multiplicity 1. As N grows to the astronomical numbers in real gases, the evenly split macrostate so utterly dominates that the system is essentially never found anywhere else.

Ludwig Boltzmann found the bridge between this counting and the thermodynamic entropy of §38.1. It is one of the most important equations in physics, engraved on his tombstone:

Key Equation: Boltzmann's entropy

$$S = k_B \ln W, \tag{38.3}$$

where $k_B = 1.38 \times 10^{-23}\ \text{J/K}$ is Boltzmann's constant. Entropy is, quite literally, a measure of the number of microscopic arrangements available to a system. And now the second law writes itself: a system left to itself drifts toward the macrostate of largest multiplicity simply because that macrostate has the most ways of happening. The entropy increases not because some force pushes it up, but because larger-W states are overwhelmingly more probable. *There is no physical law driving the increase, only the weight of numbers.*

Irreversible Processes

The distinction between reversible and irreversible processes is central, and equation (38.1) applies, strictly, only to the reversible kind.

A **reversible process** is one that can be run backward without leaving any trace on the surroundings. It requires that the system stay infinitesimally close to equilibrium at every instant: pressures and temperatures changed by vanishingly small steps, infinitely slowly. No real process is truly reversible; it is an idealization, approached by very slow, gentle changes. An **irreversible process** is everything else: everything real. A gas rushing into a vacuum, heat flowing across a finite temperature difference, a ball bouncing to rest: none can be undone without changing the wider world.

Here is the difficulty, and its elegant resolution. We can only *compute* entropy change with $\int dQ_r/T$ along a reversible path, yet the processes we care about are irreversible. The rescue is that *entropy is a state function*: its change depends only on the initial and final states, not on the path between them. So to find the entropy change of an irreversible process, we invent *any* convenient reversible path with the same endpoints, compute $\int dQ_r/T$ along that imaginary path, and the answer is the entropy change of the real process too.

Changes in Entropy

Let us put the method to work.

Example 38.1: Entropy of melting

A solid of latent heat of fusion L_f melts at temperature T_m.

a. Find the entropy change when a mass m melts.
b. Estimate it for a small ice cube.

Solution

(a) Melting happens at the constant temperature T_m, so T comes out of the integral and the heat is just $Q = mL_f$:

$$\Delta S = \int \frac{dQ_r}{T} = \frac{1}{T_m}\int dQ = \frac{Q}{T_m} = \boxed{\frac{mL_f}{T_m}}. \tag{38.1.1}$$

(b) For a $25\ \text{g}$ ice cube, with $L_f = 3.33 \times 10^5\ \text{J/kg}$ and $T_m = 273\ \text{K}$,

$$\Delta S = \frac{(0.025)(3.33 \times 10^5)}{273} \approx \boxed{30\,\mathrm{J/K}}. \tag{38.1.2}$$

The entropy rises, as it must: a disordered puddle of liquid has many more available microstates than the orderly crystal of ice.

Example 38.2: Entropy from mixing water

$1.00\ \mathrm{kg}$ of water at $0.00^\circ\mathrm{C}$ is mixed with an equal mass at $100^\circ\mathrm{C}$. The mixture settles at $50^\circ\mathrm{C}$. Find the total entropy change. (Water $c = 4186\ \mathrm{J/(kg \cdot K)}$.)

Solution

The mixing is irreversible, so we route each mass along a reversible warming or cooling path, where $dQ_r = mc\ dT$ and thus $\Delta S = \int mc\ dT/T = mc \ln(T_f/T_i)$. Working in kelvin ($T_{\mathrm{cold}} = 273$, $T_{\mathrm{hot}} = 373$, $T_f = 323$):

$$\Delta S_{\mathrm{cold}} = mc \ln \frac{T_f}{T_{\mathrm{cold}}} = (1.00)(4186) \ln \frac{323}{273} = +704\,\mathrm{J/K}, \tag{38.1.3}$$

$$\Delta S_{\mathrm{hot}} = mc \ln \frac{T_f}{T_{\mathrm{hot}}} = (1.00)(4186) \ln \frac{323}{373} = -602\,\mathrm{J/K}. \tag{38.1.4}$$

The total is

$$\Delta S = 704 - 602 = \boxed{+102\,\mathrm{J/K}}. \tag{38.1.5}$$

The cold water's entropy rises more than the hot water's falls, so the net change is positive, exactly what the second law demands for a spontaneous, irreversible process. Notice the asymmetry traces back to the T in the denominator: the cold water gains its heat at lower temperatures, where each joule buys more entropy.

Example 38.3: Free expansion: two views agree

An ideal gas undergoes an adiabatic *free expansion* to four times its initial volume. (In a free expansion the gas rushes into a vacuum, doing no work and exchanging no heat, so its temperature is unchanged.)

a. Find the entropy change by the macroscopic method.
b. Find it by the statistical method, and show the two agree.

Solution

(a) The free expansion is irreversible, so we replace it with a reversible *isothermal* expansion between the same two states (same T, volume $V_i \to 4V_i$). For an isothermal ideal-gas process $\Delta E_{\mathrm{int}} = 0$, so the heat added reversibly equals the work done by the gas, $Q_r = nRT \ln(V_f/V_i)$. Then

$$\Delta S = \frac{Q_r}{T} = nR \ln \frac{V_f}{V_i} = \boxed{nR \ln 4}. \tag{38.1.6}$$

(b) Statistically, each molecule independently has 4 times as much volume to occupy, so the multiplicity grows by a factor 4 per molecule: $W_f/W_i = 4^N$. By Boltzmann's relation,

$$\Delta S = k_B \ln \frac{W_f}{W_i} = k_B \ln\left(4^N\right) = N k_B \ln 4 = nR \ln 4, \tag{38.1.7}$$

using $Nk_B = nR$. The macroscopic and microscopic routes give $\boxed{\text{the same answer}}$: a striking confirmation that thermodynamic entropy really is Boltzmann's count of microstates.

Checkpoint 38.1

A $1.0\ \text{kg}$ block of ice melts completely at $0°\text{C}$, absorbing $3.33 \times 10^5\ \text{J}$ of heat. (a) What is its change in entropy? (b) Why is this one of the few cases where the integral in equation (38.1) can be done without calculus? (c) Does the ice's entropy increase or decrease, and why does that make sense microscopically?

Answer

Answers to Checkpoints

Checkpoint 38.1. (a) The melting happens at a constant $T = 273\ \text{K}$, so $\Delta S = Q/T = (3.33 \times 10^5\ \text{J})/(273\ \text{K}) \approx 1.22 \times 10^3\ \text{J/K}$. (b) Because the temperature stays fixed throughout the phase change, $1/T$ is a constant that comes outside the integral, leaving $\int dQ_r = Q$. (Melting at the melting point is also reversible, as the definition requires.) (c) It increases, as it must with heat flowing in. In the solid, each molecule is locked into a place in the crystal; in the liquid the molecules can wander and arrange themselves in vastly more ways, so the multiplicity, and with it $S = k_B \ln W$, rises.

This page titled 38.1: Entropy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38.2: The Second Law of Thermodynamics

We can now state the law that governs the direction of all natural processes.

Key Equation: The second law of thermodynamics

$$\Delta S \geq 0. \tag{38.4}$$

In words: *if a process occurs in an isolated system, the entropy of the system never decreases; it increases for irreversible (real) processes and stays constant for reversible (ideal) ones.* This is why heat flows from hot to cold (that direction raises total entropy), why gases mix and never spontaneously unmix, and why the bouncing ball comes to rest but never the reverse. The statistical view makes the law almost obvious: an isolated system evolves to its most probable macrostate, the one of largest multiplicity, and stays there.

Now we can answer the chapter's opening puzzle. The bouncing ball's organized, large-scale kinetic energy gradually scatters into the random thermal motion of countless molecules in the ball and floor. That scattering raises the multiplicity enormously, so it happens readily. The reverse (all those randomly jiggling molecules conspiring to push in the same direction at the same instant and launch the ball) is not forbidden by energy conservation, but its multiplicity is so vanishingly small that you would wait many times the age of the universe to see it. The **arrow of time** is the direction of increasing entropy.

Checkpoint 38.2

(a) A refrigerator makes the food inside colder, lowering its entropy. Does this violate the second law? (b) A quantity of heat Q flows by conduction from a reservoir at temperature T_h to one at $T_c < T_h$. Show that the total entropy change of the two reservoirs is positive.

Answer

Answers to Checkpoints

Checkpoint 38.2. (a) No. The refrigerator is not an isolated system: it dumps heat into the kitchen, and that heat is larger than the heat it removes from the food (it includes the work done by the motor). The entropy gained by the kitchen more than makes up for the entropy lost by the food, so the total entropy of food plus kitchen increases. (b) The hot reservoir loses heat Q at temperature T_h, so $\Delta S_h = -Q/T_h$; the cold one gains Q at T_c, so $\Delta S_c = +Q/T_c$. The total is $\Delta S = Q(1/T_c - 1/T_h)$, which is positive because $T_c < T_h$ makes $1/T_c > 1/T_h$. That is why heat flows spontaneously from hot to cold and never the reverse.

This page titled 38.2: The Second Law of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38.3: Heat Engines

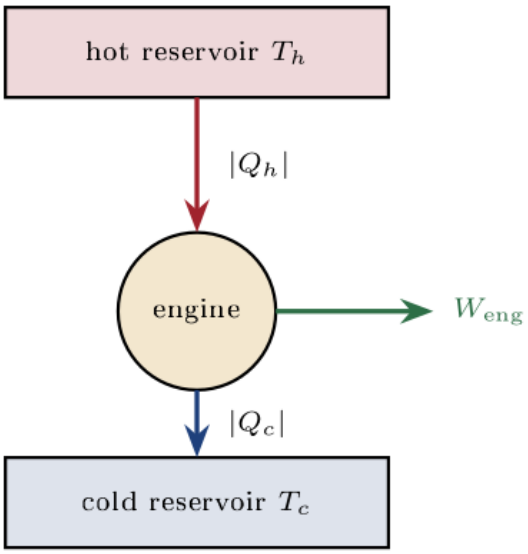


Figure 38.2. A heat engine. Each cycle it takes in heat $|Q_h|$ from the hot reservoir, converts part of it to work W_{eng}, and discards the rest as waste heat $|Q_c|$ to the cold reservoir. The second law forbids converting *all* of $|Q_h|$ to work.

•Figure description

An energy-flow diagram of a heat engine. A red block at the top is the hot reservoir at temperature T sub h; a blue block at the bottom is the cold reservoir at temperature T sub c; a gold circle between them is the engine. A thick arrow carries heat of magnitude Q sub h down from the hot reservoir into the engine. A smaller arrow carries waste heat of magnitude Q sub c on down from the engine into the cold reservoir. A third arrow leaves the engine sideways to the right as the work output W sub eng. The width of the flows shows that only part of the heat taken in becomes work.

A **heat engine** turns heat into useful work by running a working substance around a **cyclic process**: a closed loop that returns to its starting state each cycle. Because entropy and internal energy are state functions, both return to their starting values every cycle, so $\Delta S = 0$ and $\Delta E_{\text{int}} = 0$ over a complete cycle. An engine draws heat Q_h from a hot reservoir, dumps waste heat Q_c to a cold reservoir, and delivers the difference as work (Figure 38.2):

Key Equation: Work of a heat engine

$$W_{\text{eng}} = |Q_h| - |Q_c|. \tag{38.5}$$

The **efficiency** e is what you get out divided by what you put in: work delivered per unit heat drawn from the hot reservoir:

Key Equation: Thermal efficiency of a heat engine

$$e = \frac{W_{\text{eng}}}{|Q_h|} = \frac{|Q_h| - |Q_c|}{|Q_h|} = 1 - \frac{|Q_c|}{|Q_h|}. \tag{38.6}$$

An efficiency of $1\ (100\%)$ would mean no waste heat at all, $|Q_c| = 0$, every joule of heat turned into work. The second law forbids it: such an engine would lower the entropy of the universe. Some heat must always be thrown away.

Engine Cycles on a P–V Diagram

In Chapter 37 we saw that the work a gas does during a process is the area under its path on a P–V diagram, $W_{\text{by}} = \int P\ dV$: positive when the gas expands and negative when it is compressed. An engine's working substance goes around a closed loop, so in each cycle it expands along one part of the loop and is compressed along another. If the expansion happens at high pressure

and the compression at low pressure, the gas does more work expanding than is done on it during compression. The net work is the area under the upper path minus the area under the lower path, which is simply the area *enclosed* by the loop (Figure 38.3):

Key Equation: Net work in a cycle

$$W_{\text{eng}} = W_{\text{by, net}} = \text{area enclosed by the cycle on the } P\text{--}V \text{ diagram.} \tag{38.7}$$

The direction matters. A loop traversed *clockwise* expands at high pressure and is compressed at low pressure, so the gas does positive net work: it is an engine. A loop traversed *counterclockwise* has net work done *on* the gas: it is a refrigerator or heat pump (Section 38.3). In either case $\Delta E_{\text{int}} = 0$ over the complete cycle, so the first law says the net heat absorbed equals the net work done by the gas, $Q_{\text{net}} = W_{\text{by, net}}$.

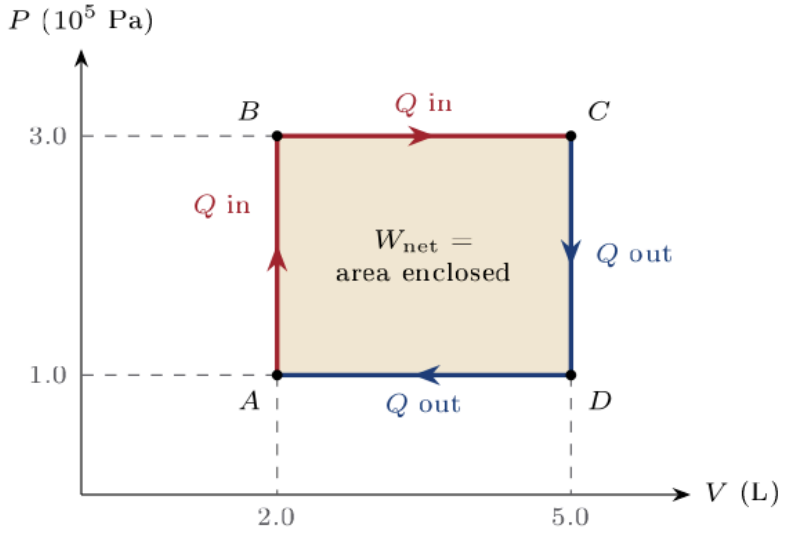


Figure 38.3. A rectangular engine cycle, traversed clockwise. The gas expands at high pressure ($B \to C$) and is compressed at low pressure ($D \to A$), so the net work it does is the shaded area enclosed by the loop. Heat enters on the red legs, where the gas warms; heat leaves on the blue legs, where it cools. Example 38.4 works through this cycle.

•Figure description

A pressure-versus-volume diagram of a rectangular cycle traversed clockwise through corners A, B, C and D. A is at the lower left, at 2.0 liters and 1.0 times 10 to the 5 pascals. From A a vertical line rises to B at 3.0 times 10 to the 5 pascals, then a horizontal line runs right to C at 5.0 liters, a vertical line drops to D at 1.0 times 10 to the 5 pascals, and a horizontal line runs left back to A. The enclosed rectangle is shaded and labeled W net equals area enclosed. Red labels mark heat flowing in on the two legs A to B and B to C; blue labels mark heat flowing out on the legs C to D and D to A.

To find an engine's efficiency from its diagram we also need $|Q_h|$, the heat taken in. This is the sum of Q over only those legs on which heat flows *into* the gas; the legs on which heat flows out make up $|Q_c|$. The first law gives the heat on each leg as $Q = \Delta E_{\text{int}} - W = \Delta E_{\text{int}} + W_{\text{by}}$. For a monatomic ideal gas, $E_{\text{int}} = \frac{3}{2}nRT = \frac{3}{2}PV$, so the internal-energy change can be read straight off the diagram as $\Delta E_{\text{int}} = \frac{3}{2}\,\Delta(PV)$.

Example 38.4: Efficiency of a rectangular cycle

A monatomic ideal gas is taken around the clockwise cycle of Figure 38.3: A (2.0 L, 1.0×10^5 Pa) $\to$ B (2.0 L, 3.0×10^5 Pa) $\to$ C (5.0 L, 3.0×10^5 Pa) $\to$ D (5.0 L, 1.0×10^5 Pa) $\to$ A.

a. Find the net work done by the gas per cycle.
b. Find the heat absorbed per cycle, $|Q_h|$, and the efficiency.

Solution

Work in SI units: $1.0\ \text{L} = 1.0 \times 10^{-3}\ \text{m}^3$, so the products PV at the corners are 200 J at A, 600 J at B, 1500 J at C, and 500 J at D.

(a) The loop is a rectangle, so the enclosed area is its width times its height:

$$W_{\text{eng}} = (\Delta P)(\Delta V) = (2.0 \times 10^5\ \text{Pa})(3.0 \times 10^{-3}\ \text{m}^3) = \boxed{600\ \text{J}}. \tag{38.3.1}$$

(b) Heat flows in on the two legs where PV, and so the temperature, rises.

$A \to B$ (isovolumetric): no work, so $Q = \Delta E_{\text{int}} = \frac{3}{2}\ \Delta(PV) = \frac{3}{2}(600 - 200)\ \text{J} = 600\ \text{J}$.

$B \to C$ (isobaric): $W_{\text{by}} = P\ \Delta V = (3.0 \times 10^5)(3.0 \times 10^{-3}) = 900\ \text{J}$ and $\Delta E_{\text{int}} = \frac{3}{2}(1500 - 600)\ \text{J} = 1350\ \text{J}$, so $Q = 1350 + 900 = 2250\ \text{J}$.

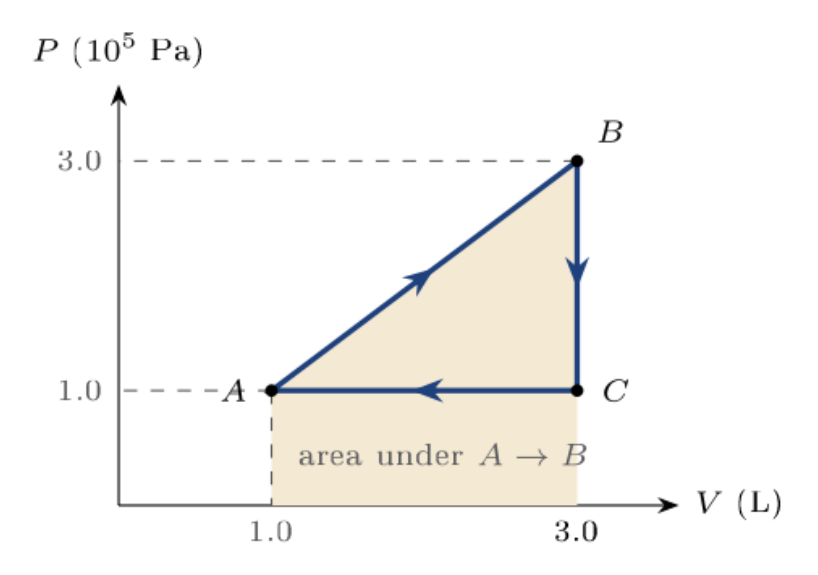
Together these give $|Q_h| = 600 + 2250 = 2850\ \text{J}$, and

$$e = \frac{W_{\text{eng}}}{|Q_h|} = \frac{600\ \text{J}}{2850\ \text{J}} = \boxed{0.21 = 21\%}. \tag{38.3.2}$$

(The other two legs give $Q = -1500\ \text{J}$ and $-750\ \text{J}$, so $|Q_c| = 2250\ \text{J}$, and $|Q_h| - |Q_c| = 600\ \text{J}$, as it must.)

Example 38.5: A cycle with a sloping leg

A monatomic ideal gas runs the clockwise triangular cycle shown: from A (1.0 L, 1.0×10^5 Pa) along a straight line to B (3.0 L, 3.0×10^5 Pa), straight down at constant volume to C (3.0 L, 1.0×10^5 Pa), and back to A at constant pressure.



•Figure description
A pressure-versus-volume graph of a triangular cycle traversed clockwise. Point A is at 1.0 liter and 1.0 times 10 to the 5 pascals. A straight sloping line runs up and to the right from A to point B at 3.0 liters and 3.0 times 10 to the 5 pascals. A vertical line runs down from B to point C at 3.0 liters and 1.0 times 10 to the 5 pascals, and a horizontal line runs left from C back to A. The area under the sloping line, down to the volume axis, is lightly shaded as a trapezoid.

a. Find the net work done by the gas per cycle.
b. Find Q for each leg, and the efficiency.

Solution

The products PV at the corners are 100 J at A, 900 J at B, and 300 J at C.

(a) The enclosed triangle has base $\Delta V = 2.0 \times 10^{-3}\ \text{m}^3$ and height $\Delta P = 2.0 \times 10^5\ \text{Pa}$:

$$W_{\text{eng}} = \tfrac{1}{2}(2.0 \times 10^{-3}\ \text{m}^3)(2.0 \times 10^5\ \text{Pa}) = \boxed{200\ \text{J}}. \tag{38.3.3}$$

(b) On each leg use W_{by} = area under the path and $\Delta E_{\text{int}} = \frac{3}{2}\ \Delta(PV)$, then $Q = \Delta E_{\text{int}} + W_{\text{by}}$.

$A \to B$ (sloping line): this leg has *both* work and a change in internal energy. The area under a straight line is a trapezoid, whose height is the average pressure:

$$\begin{aligned} W_{\text{by}} &= \tfrac{1}{2}(1.0 + 3.0) \times 10^5\,\text{Pa} \times 2.0 \times 10^{-3}\,\text{m}^3 = 400\,\text{J}, \\ \Delta E_{\text{int}} &= \tfrac{3}{2}(900 - 100)\,\text{J} = 1200\,\text{J}, \end{aligned} \tag{38.3.4}$$

so $Q_{AB} = 1200 + 400 = +1600\ \text{J}$.

$B \to C$ (constant volume): $W_{\text{by}} = 0$ and $\Delta E_{\text{int}} = \frac{3}{2}(300 - 900)\ \text{J} = -900\ \text{J}$, so $Q_{BC} = -900\ \text{J}$.

$C \to A$ (constant pressure, compression): $W_{\text{by}} = (1.0 \times 10^5)(-2.0 \times 10^{-3}) = -200\ \text{J}$ and $\Delta E_{\text{int}} = \frac{3}{2}(100 - 300)\ \text{J} = -300\ \text{J}$, so $Q_{CA} = -500\ \text{J}$.

Heat enters only on $A \to B$, so $|Q_h| = 1600\ \text{J}$ and

$$e = \frac{W_{\text{eng}}}{|Q_h|} = \frac{200\,\text{J}}{1600\,\text{J}} = \boxed{0.125 = 12.5\%}. \tag{38.3.5}$$

Check: the three heats add to $1600 - 900 - 500 = 200\ \text{J}$, the net work, as the first law requires over a full cycle. Notice that most of the heat absorbed on the sloping leg ($1200\ \text{J}$ of the $1600\ \text{J}$) goes into raising the gas's internal energy, which it must then give back on the other two legs; only the enclosed area is left over as work.

Efficiency

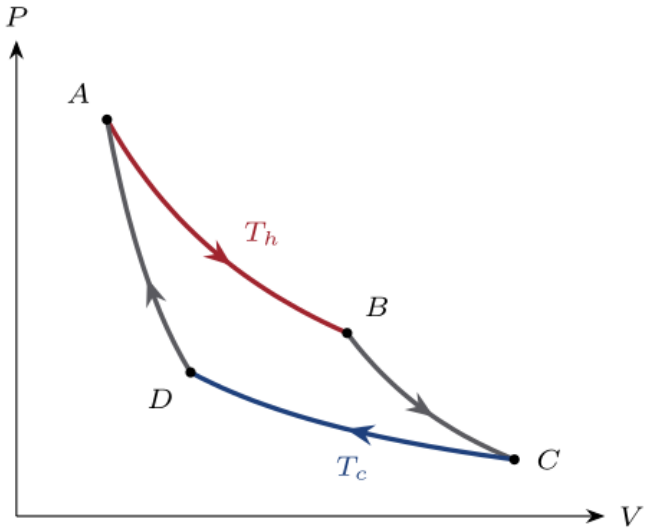


Figure 38.4. The Carnot cycle: two isothermal steps (red at T_h, blue at T_c) joined by two adiabatic steps (gray). This reversible loop is the most efficient cycle possible between the two temperatures, with efficiency $e_C = 1 - T_c/T_h$. (The adiabats are drawn steeper than they would be for a real gas, so the two kinds of curve are easy to tell apart.)

•Figure description

A pressure-versus-volume diagram of the Carnot cycle, a closed four-sided loop traversed clockwise through corners labeled A, B, C and D. From A to B a red curve is an isothermal expansion at the hot temperature T sub h. From B to C a gray curve is an adiabatic expansion, falling more steeply. From C to D a blue curve is an isothermal compression at the cold temperature T sub c. From D back to A a second gray curve is an adiabatic compression. Two isotherms joined by two adiabats, all reversible.

How efficient *can* an engine be? Sadi Carnot answered this in 1824 by imagining the most efficient possible engine: one running a reversible cycle of two isothermal and two adiabatic steps, the **Carnot cycle** (Figure 38.4). He proved a remarkable thing: that the maximum efficiency depends on *nothing* but the temperatures of the two reservoirs. Written in terms of absolute temperature (a scale Kelvin introduced a generation later, partly on the strength of Carnot's argument), it is

Key Equation: Carnot (maximum) efficiency

$$e_C = 1 - \frac{T_c}{T_h}, \tag{38.8}$$

with T_c and T_h in kelvin. No engine operating between two given temperatures can beat this; real engines, plagued by friction and irreversibility, fall short of it. The formula carries a sobering message: the only way to reach 100% efficiency would be a cold reservoir at absolute zero ($T_c = 0$), which is unattainable. To build a better engine you want the hottest possible source and the coldest possible sink.

Example 38.6: A steam engine's ceiling

A steam engine boils water at $500\ \mathrm{K}$ and exhausts to the outside air at about $300\ \mathrm{K}$. What is its maximum possible thermal efficiency?

Solution

The best any engine can do between these reservoirs is the Carnot efficiency,

$$e_C = 1 - \frac{T_c}{T_h} = 1 - \frac{300\,\mathrm{K}}{500\,\mathrm{K}} = 1 - 0.60 = \boxed{0.40 = 40\%}. \tag{38.3.6}$$

Even a flawless, frictionless engine here must dump 60% of its intake heat as waste: a hard limit set by the temperatures alone, not by engineering skill.

Example 38.7: Pushing the hot side higher

The hottest temperature standard engine materials can withstand is about $850\ \mathrm{K}$. With the atmosphere as the cold reservoir at $300\ \mathrm{K}$, what is the highest theoretical efficiency?

Solution

$$e_C = 1 - \frac{T_c}{T_h} = 1 - \frac{300\,\mathrm{K}}{850\,\mathrm{K}} = 1 - 0.353 \approx \boxed{0.65 = 65\%}. \tag{38.3.7}$$

Raising the hot reservoir from $500\ \mathrm{K}$ to $850\ \mathrm{K}$ lifted the ceiling from 40% to 65%, which is exactly why engineers chase ever-higher operating temperatures, limited in the end by what the materials can survive.

The Carnot limit also explains why the cycles of Section 38.3 perform so poorly. In the rectangular cycle of Example 38.4, the gas is coldest at A and hottest at C, and since $T = PV/nR$, $T_C/T_A = 1500/200 = 7.5$. A Carnot engine working between those two temperatures would reach $e_C = 1 - 1/7.5 = 87\%$, yet the rectangle manages only 21%. The reason is *where* the heat flows. A Carnot engine takes in all of its heat at the hottest temperature and rejects all of it at the coldest. The rectangle takes in heat all the way up from T_A to T_C and rejects it all the way back down, so on average it absorbs heat well below its top temperature and discards it well above its bottom one, and every one of those heat flows across a temperature difference is irreversible.

Refrigerators

Run a heat engine backward and you get a **refrigerator** (or a **heat pump**). Instead of letting heat flow hot-to-cold and harvesting work, a refrigerator uses work input to *force* heat from a cold region to a hot one: the unnatural direction. This does not violate the second law: the work you supply (ultimately from the wall outlet) generates enough extra entropy to more than pay for the entropy decrease of moving heat the "wrong" way.

We rate a refrigerator not by efficiency but by its **coefficient of performance** (COP), the heat extracted from the cold space per unit work supplied:

Key Equation: Refrigerator coefficient of performance

$$K = \frac{|Q_c|}{W} = \frac{|Q_c|}{|Q_h| - |Q_c|}. \tag{38.9}$$

A good refrigerator has $K > 1$: it moves more heat than the work it consumes, because it is pumping existing heat rather than creating it. As with engines, the second law sets a best-possible value, achieved by a reversible (Carnot) refrigerator,

Key Equation: Carnot (maximum) coefficient of performance

$$K_C = \frac{T_c}{T_h - T_c}, \tag{38.10}$$

which shows that a refrigerator works hardest, and its COP plummets, when it must maintain a large temperature difference between its cold interior and the warm room outside.

Checkpoint 38.3

(a) What is the maximum efficiency of an engine operating between reservoirs at $500\ \mathrm{K}$ and $300\ \mathrm{K}$? (b) Another engine operates between $600\ \mathrm{K}$ and $400\ \mathrm{K}$: the same $200\ \mathrm{K}$ difference. Is its maximum efficiency the same? (c) Why can no engine have an efficiency of 1?

Answer

Answers to Checkpoints

Checkpoint 38.3. (a) The Carnot efficiency is $e_C = 1 - T_c/T_h = 1 - 300/500 = 0.40$, or 40%. (b) No: $e_C = 1 - 400/600 = 0.33$, lower. The efficiency depends on the *ratio* of the temperatures, not their difference, and the same $200\ \mathrm{K}$ is a smaller fraction of the hotter temperature. (c) An efficiency of 1 would require $|Q_c| = 0$ (all the heat turned into work) or $T_c = 0$. Dumping no heat would decrease the total entropy, since heat taken from the hot reservoir lowers its entropy and nothing compensates, violating the second law. And a cold reservoir at absolute zero is unattainable.

This page titled 38.3: Heat Engines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38.S: Entropy and the Second Law of Thermodynamics (Summary)

Summary

- **Entropy** is a state function with $dS = dQ_r/T$ (reversible); $\Delta S = \int dQ_r/T$. Same heat $\rightarrow$ more entropy at lower T.
- **Statistical view**: a *macrostate* has multiplicity W (number of microstates); for a divided box $W = N!/(n_1!n_2!)$. **Boltzmann**: $S = k_B \ln W$. Systems drift to the largest-W macrostate purely by probability.
- **Reversible** processes are infinitely slow idealizations; real processes are **irreversible**. Because S is a state function, compute an irreversible process's ΔS along any reversible path with the same endpoints (e.g. $\Delta S = mc\ln(T_f/T_i)$ for heating; $nR\ln(V_f/V_i)$ for isothermal expansion).
- **Second law**: $\Delta S \geq 0$ for an isolated system; entropy never decreases. It fixes the arrow of time and the direction of heat flow.
- **Heat engines** (cyclic): $W_{\text{eng}} = |Q_h| - |Q_c|$ and efficiency $e = 1 - |Q_c|/|Q_h|$. On a P–V diagram the net work is the **area enclosed** by the cycle (clockwise for an engine), and $|Q_h|$ is the sum of Q over the legs where heat flows in. The maximum is the **Carnot efficiency** $e_C = 1 - T_c/T_h$ (T in kelvin); 100% is impossible.
- **Refrigerators** run engines in reverse, rated by the coefficient of performance $K = |Q_c|/W$, with Carnot limit $K_C = T_c/(T_h - T_c)$.

Key Equations

Entropy change (reversible)

$$dS = \frac{dQ_r}{T} \qquad \Longrightarrow \qquad \Delta S = \int \frac{dQ_r}{T} \tag{38.S.1}$$

Multiplicity of a divided box

$$W = \frac{N!}{n_1!\, n_2!} \tag{38.S.2}$$

Boltzmann's entropy

$$S = k_B \ln W \tag{38.S.3}$$

Entropy of heating, and of isothermal expansion

$$\begin{aligned} \Delta S &= mc \ln \frac{T_f}{T_i} \\ \Delta S &= nR \ln \frac{V_f}{V_i} \end{aligned} \tag{38.S.4}$$

The second law of thermodynamics

$$\Delta S \geq 0 \tag{38.S.5}$$

Work of a heat engine

$$W_{\text{eng}} = |Q_h| - |Q_c| \tag{38.S.6}$$

Net work in a cycle

$$W_{\text{eng}} = \text{area enclosed on the } P\text{--}V \text{ diagram} \tag{38.S.7}$$

Thermal efficiency

$$e = \frac{W_{\text{eng}}}{|Q_h|} = 1 - \frac{|Q_c|}{|Q_h|} \tag{38.S.8}$$

Carnot (maximum) efficiency

$$e_C = 1 - \frac{T_c}{T_h} \qquad (T \text{ in kelvin}) \tag{38.S.9}$$

Refrigerator coefficient of performance

$$K = \frac{|Q_c|}{W} = \frac{|Q_c|}{|Q_h| - |Q_c|} \tag{38.S.10}$$

Carnot (maximum) coefficient of performance

$$K_C = \frac{T_c}{T_h - T_c} \tag{38.S.11}$$

This page titled 38.S: Entropy and the Second Law of Thermodynamics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

38.E: Entropy and the Second Law of Thermodynamics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $c_{\text{water}} = 4186\ \text{J/(kg·K)}$, $L_f = 3.33 \times 10^5\ \text{J/kg}$, $R = 8.31\ \text{J/(mol·K)}$, $k_B = 1.38 \times 10^{-23}\ \text{J/K}$, and temperatures in kelvin.

Section 38.1 — Entropy and the Statistical View

38.1 ★ A reservoir at $400\ \text{K}$ absorbs $2000\ \text{J}$ of heat reversibly. Find its entropy change.

38.2 ★★ Four particles are in a divided box.

a. Find the multiplicity of the macrostate with two particles on each side.
b. Find the multiplicity of the "all four on the left" macrostate.
c. Which macrostate is more likely, and by what factor?

38.3 ★★ *[Conceptual.]* Using $S = k_B \ln W$, explain why a shuffled deck of cards has higher entropy than a sorted one, and why the second law makes spontaneous "unshuffling" essentially impossible.

38.4 ★ A large reservoir at $300\ \text{K}$ rejects $1500\ \text{J}$ of heat reversibly. Find its entropy change, and state its sign.

38.5 ★★★ Eight identical gas molecules rattle around a box divided into two equal halves.

a. How many microstates correspond to the macrostate "five on the left, three on the right"?
b. How many correspond to "four and four"?
c. Using $S = k_B \ln W$, find the entropy difference between the two macrostates and state which is more probable.

Sections 38.1.2–38.1.3 — Irreversible Processes and Entropy Changes

38.6 ★★ Find the entropy change when $0.50\ \text{kg}$ of ice melts at 0°C.

38.7 ★★ $2.0\ \text{kg}$ of water is heated reversibly from 20°C to 80°C. Find the entropy change. *(Use $\Delta S = mc\ln(T_f/T_i)$ with T in kelvin.)*

38.8 ★★★ $3.0\ \text{mol}$ of an ideal gas expands isothermally to three times its initial volume.

a. Find the entropy change by the macroscopic method.
b. Confirm it with the statistical method ($W \propto V^N$).

38.9 ★★ A $1.5\ \text{kg}$ block of copper ($c = 386\ \text{J/(kg·K)}$) is heated reversibly from 20°C to 120°C. Find its entropy change.

38.10 ★★★ $2.0\ \text{mol}$ of an ideal gas expands to four times its initial volume.

a. Find the entropy change for a reversible isothermal expansion.
b. A free expansion into vacuum produces the same change of state. Is the gas's entropy change the same? Explain why a state function does not depend on the path.

Section 38.2 — The Second Law

38.11 ★ *[Conceptual.]* For each, state whether the system's entropy increases, decreases, or stays the same, and reconcile it with the second law: (a) water freezing into ice in a freezer; (b) a gas expanding freely into a vacuum; (c) a reversible adiabatic expansion.

38.12 ★★ $500\ \text{g}$ of water at 80°C is poured into a large lake at 20°C. Estimate the total entropy change of water-plus-lake, treating the lake as a reservoir at constant 20°C, and confirm it is positive.

38.13 ★ *[Conceptual.]* An inventor claims a device that each cycle draws $500\ \text{J}$ from a hot reservoir and converts *all* of it into work, rejecting no heat. Which statement of the second law does this violate, and why is it impossible?

38.14 ★★ Heat flows steadily through a metal bar whose ends are held in contact with a hot reservoir at $600\ \text{K}$ and a cold reservoir at $300\ \text{K}$. During one interval, $1200\ \text{J}$ passes through the bar.

a. Find the entropy change of each reservoir.
b. Find the total entropy change, and show that it is positive, as the second law requires.

Section 38.3 — Engines, Efficiency, Refrigerators & the Arrow of Time

38.15 ★ An engine takes in $800\ \text{J}$ of heat per cycle and exhausts $500\ \text{J}$. Find (a) the work done per cycle and (b) the efficiency.

38.16 ★★ A Carnot engine operates between reservoirs at $600\ \text{K}$ and $300\ \text{K}$.

a. Find its efficiency.
b. If it absorbs $1200\ \text{J}$ from the hot reservoir per cycle, how much work does it deliver and how much heat does it reject?

38.17 ★★ A refrigerator extracts $300\ \text{J}$ of heat from its cold interior while its motor does $120\ \text{J}$ of work each cycle.

a. Find the coefficient of performance.
b. How much heat is dumped into the kitchen per cycle?

38.18 ★★ *[Real-world application: engineering.]* A steam power plant runs between a boiler at $810\ \text{K}$ and a river at $300\ \text{K}$.

a. Find the maximum possible (Carnot) efficiency.
b. Operating at this ideal efficiency while delivering $600\ \text{MW}$ of electrical power, at what rate must it dump waste heat into the river?

38.19 ★★★ A heat pump keeps a house at $295\ \text{K}$ by drawing heat from outdoor air at $275\ \text{K}$. For heating, the coefficient of performance is $K_{\text{hp}} = |Q_h|/W$.

a. Show that its maximum (Carnot) value is $K_{\text{hp}} = T_h/(T_h - T_c)$, and evaluate it.
b. How much heat is delivered to the house per joule of electrical work at this limit?
c. Explain why this beats an electric resistance heater, which delivers exactly $1\ \text{J}$ of heat per joule of work.

38.20 ★★★ *[Conceptual — the arrow of time.]* Reread the chapter opener about the bouncing ball.

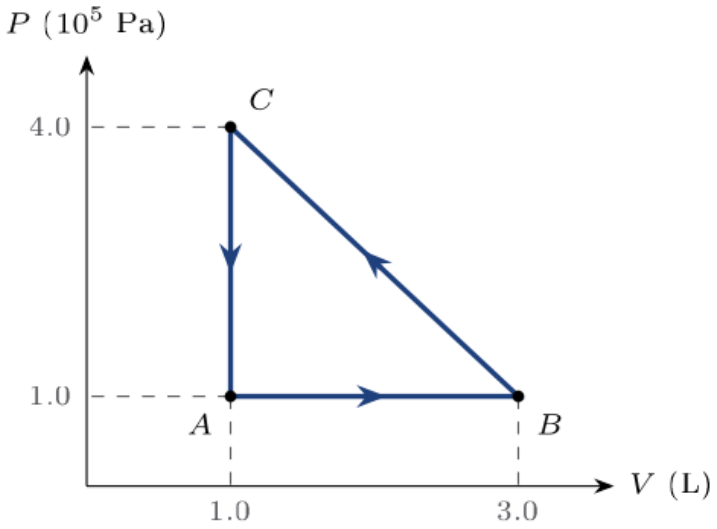
a. A bouncing ball comes to rest, its kinetic energy spreading into thermal motion of the ball and floor. Explain, in terms of multiplicity, why this happens readily but the reverse does not.
b. Energy would still be conserved if a resting ball spontaneously began to bounce. Which law forbids it, and what does this reveal about why time seems to flow in only one direction?
c. Does the second law make the spontaneous re-bouncing strictly *impossible*, or merely overwhelmingly improbable? Briefly justify your answer.

Section 38.3.1 — Engine Cycles on a P–V Diagram

38.21 ★ An engine's working gas is taken clockwise around a cycle that encloses an area of $450\ \text{J}$ on its P–V diagram. Each cycle the gas absorbs $1800\ \text{J}$ of heat.

a. How much net work does the gas do per cycle?
b. What is the engine's efficiency?
c. How much heat does it exhaust per cycle?

38.22 ★★ *[Graph reading.]* A gas is taken around the triangular cycle $A \to B \to C \to A$ shown.

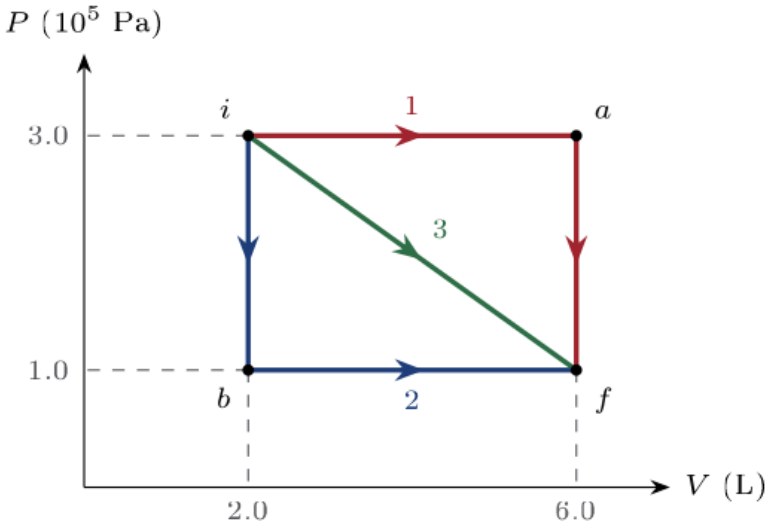


•Figure description

A pressure-versus-volume graph of a triangular cycle. Point A is at 1.0 liter and 1.0 times 10 to the 5 pascals. A horizontal line runs right from A to point B at 3.0 liters and the same pressure. A straight diagonal line runs up and to the left from B to point C at 1.0 liter and 4.0 times 10 to the 5 pascals. A vertical line runs down from C back to A. Arrows on each side show the direction A to B to C to A.

a. What is the magnitude of the net work per cycle?
b. Is the net work done *by* the gas or *on* it? How can you tell from the graph?
c. What is the net heat absorbed by the gas per cycle (with sign)?
d. Could this cycle be the cycle of a heat engine? Explain.

38.23 ★★ *[Graph reading — work depends on the path.]* An ideal gas is taken from state i ($2.0\ \text{L}$, $3.0 \times 10^5\ \text{Pa}$) to state f ($6.0\ \text{L}$, $1.0 \times 10^5\ \text{Pa}$) along each of the three paths shown: path 1 through a, path 2 through b, and path 3 straight from i to f.



•Figure description

A pressure-versus-volume graph with an initial point i at 2.0 liters and 3.0 times 10 to the 5 pascals and a final point f at 6.0 liters and 1.0 times 10 to the 5 pascals. Path 1 runs horizontally right from i to point a at 6.0 liters and 3.0 times 10 to the 5 pascals, then straight down to f. Path 2 runs straight down from i to point b at 2.0 liters and 1.0 times 10 to the 5 pascals, then horizontally right to f. Path 3 is a straight diagonal line from i down to f.

a. Find the work done by the gas along each path.

b. Show that the gas has the same temperature at i and at f. What, then, is ΔE_{int} along each path?
c. Find the heat absorbed along each path. Which quantities in this problem depend on the path, and which do not?

38.24 ★★★ $0.080\ \text{mol}$ of a monatomic ideal gas runs the cycle A ($2.0\ \text{L}$, $1.0\times10^5\ \text{Pa}$) $\to$ B ($2.0\ \text{L}$, $2.0\times10^5\ \text{Pa}$) $\to$ C ($4.0\ \text{L}$, $2.0\times10^5\ \text{Pa}$) $\to$ D ($4.0\ \text{L}$, $1.0\times10^5\ \text{Pa}$) $\to$ A.

a. Sketch the cycle on a P–V diagram and state whether it runs clockwise or counterclockwise. Find the temperature at each corner.
b. For each leg, find the work done by the gas, the change in internal energy, and the heat absorbed.
c. Find the net work per cycle, $|Q_h|$, and the efficiency.
d. Find the Carnot efficiency between the highest and lowest temperatures in the cycle, and explain why this engine falls so far short of it.

Challenge Problems

38.25 ★★★ *[Real-world application: energy — thermal pollution from a power plant.]* A coal-fired station delivers $1000\ \text{MW}$ of electrical power. Its boiler runs at $830\ \text{K}$ and it rejects heat to a river at $295\ \text{K}$. Its measured overall efficiency is 38%.

a. Find the Carnot efficiency for these two reservoir temperatures.
b. What fraction of the Carnot limit does the plant actually achieve?
c. At what rate does the plant dump waste heat into the river?
d. Environmental rules cap the river's temperature rise at $5.0\ \text{K}$. What minimum river flow rate is required, in kg/s and m^3/s? Compare this with a typical small river's flow of a few tens of m^3/s, and explain why large plants are built on large rivers, on coastlines, or with cooling towers.

38.26 ★★★ *[Real-world application: engineering — the price of mixing.]* $2.0\ \text{kg}$ of water at 80°C is poured into $3.0\ \text{kg}$ of water at 20°C in a perfectly insulated container. Take $c = 4186\ \text{J/(kg}\cdot\text{K)}$.

a. Find the final temperature.
b. Using $\Delta S = mc\ln(T_f/T_i)$ for each portion, find the entropy change of the hot water, of the cold water, and of the system as a whole. Confirm the sign of the total.
c. No heat left the container, so no energy was lost. Yet something *was* lost. The maximum work that could have been extracted (had the two portions been brought together through a reversible engine instead of simply poured) is $W_{\text{lost}} = T_0\ \Delta S_{\text{total}}$, with $T_0 = 293\ \text{K}$ the temperature of the surroundings. Evaluate it.
d. Explain in two or three sentences why the first law alone cannot distinguish the pouring from the reversible process, and what quantity does the distinguishing.

38.27 ★★★ *[Real-world application: refrigeration — making ice.]* A freezer holds its interior at -18°C in a kitchen at 22°C. An ice tray holding $2.0\ \text{kg}$ of water at 20°C is placed inside. Take $c_{\text{water}} = 4186\ \text{J/(kg}\cdot\text{K)}$, $c_{\text{ice}} = 2100\ \text{J/(kg}\cdot\text{K)}$, and $L_f = 3.33\times10^5\ \text{J/kg}$.

a. Find the total heat that must be removed from the tray to bring it to ice at -18°C. Identify which of the three stages dominates.
b. Find the Carnot coefficient of performance for a refrigerator working between these two temperatures.
c. What is the least possible work the compressor could do to remove that heat?
d. A real freezer of this size achieves $K = 2.2$. Find the actual work required and, if the compressor draws $150\ \text{W}$ of mechanical power, how long it must run. Then find the total entropy change of the universe: add the entropy change of the tray's contents (cooling the water, freezing it, then cooling the ice, three terms, using $\Delta S = mc\ln(T_f/T_i)$ and $\Delta S = -mL_f/T_{\text{freeze}}$) to that of the kitchen, which absorbs $|Q_h| = |Q_c| + W$ at 22°C. Confirm the sign.

This page titled 38.E: Entropy and the Second Law of Thermodynamics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39: INTRODUCTION TO QUANTUM MECHANICS

Heat a poker in a fire and it glows: first a dull red, then orange, then (hotter still) a brilliant white. Every object radiates, and the color of the glow betrays the temperature. This much classical physics could describe. But when physicists at the close of the nineteenth century tried to *calculate* the spectrum of that glow from first principles, they got an absurd answer. Their best theory predicted that any warm object should pour out an infinite amount of energy at short wavelengths: that the inside of your oven should be flooding the room with ultraviolet light and X-rays. It plainly does not. This failure, the “ultraviolet catastrophe,” was not a small discrepancy to be patched; it was a sign that something in the classical picture was deeply, structurally wrong.

The fix, found by Planck in 1900 almost against his will, was a single strange idea: that energy comes in discrete lumps, not a smooth continuum. From that one crack the whole edifice of classical physics gave way over the next quarter century. Light, long known to be a wave, turned out to also arrive as particles (*photons*) with definite energy and momentum, as the photoelectric effect and Compton scattering proved. And then, in a breathtaking symmetry, matter turned out to also be wavelike: an electron has a wavelength. Finally, Heisenberg showed that this wave–particle nature sets a hard limit on what can be known at all. This chapter follows that revolution from the glowing poker to the uncertainty principle.

Learning Objectives

- Explain blackbody radiation and the ultraviolet catastrophe, and apply Wien’s displacement law.
- Use Planck’s quantization of oscillator energy, $E = nhf$.
- Apply the photoelectric equation to find maximum kinetic energies, stopping potentials, and cutoff wavelengths.
- Treat light as photons carrying energy $E = hf$ and momentum $p = h/\lambda$.
- Compute the Compton shift in wavelength when a photon scatters from an electron.
- Find the de Broglie wavelength of a moving particle.
- Apply the Heisenberg uncertainty principle to position–momentum and energy–time.

The through-line is a single question: how much of the classical world survives when energy comes in lumps? We will return to wave–particle duality at the end.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

39.1: Blackbody Radiation
39.2: Photoelectric Effect
39.3: Photons
39.4: Compton Shift
39.5: de Broglie Wavelength
39.6: Heisenberg Uncertainty Principle
39.S: Introduction to Quantum Mechanics (Summary)
39.E: Introduction to Quantum Mechanics (Problems)

This page titled 39: Introduction to Quantum Mechanics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.1: Blackbody Radiation

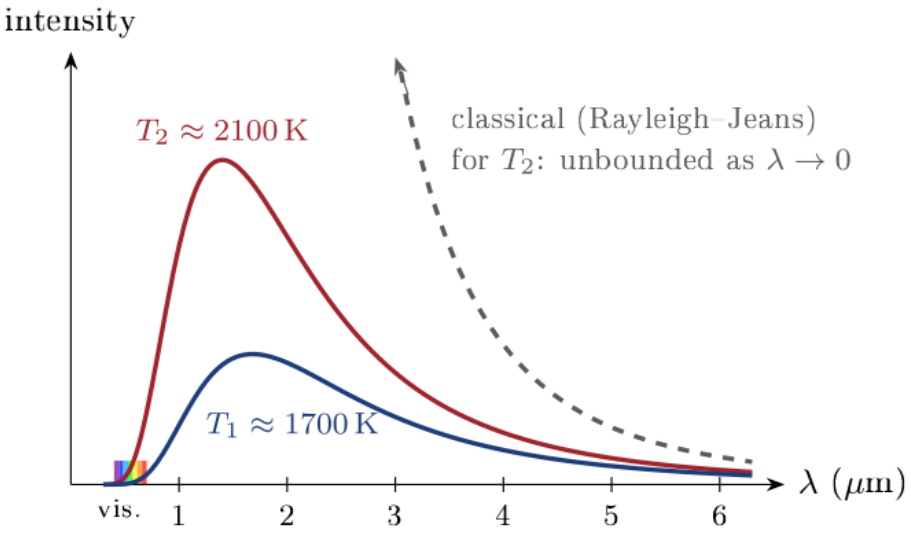


Figure 39.1. Blackbody radiation. Measured curves (solid) peak and fall off; the peak moves to shorter λ as temperature rises (Wien's law). At these temperatures, near that of a lamp filament or a poker in a fire, the peak is in the infrared and only the short-wavelength tail reaches the visible band, which is why such objects glow a dull red. The classical Rayleigh–Jeans prediction (dashed) matches at long λ but diverges as $\lambda \to 0$: the ultraviolet catastrophe.

•Figure description

A graph of emitted intensity against wavelength in micrometers, from 0 to 6, for a blackbody. A narrow rainbow-colored strip just above the axis between 0.4 and 0.7 micrometers, labelled vis., marks the visible band. Two solid curves are shown, labelled T sub 2 about 2100 kelvin and T sub 1 about 1700 kelvin, each rising from zero at short wavelength to a rounded peak and then falling away slowly at long wavelength. The hotter curve is taller and its peak, at about 1.4 micrometers in the infrared, sits at a shorter wavelength than the cooler curve's peak at about 1.7 micrometers: Wien's law. Only the short-wavelength tails of the curves reach the visible strip. A dashed gray curve, the classical Rayleigh-Jeans prediction for the hotter temperature, lies above the hotter solid curve and approaches it at long wavelength, but instead of turning over it climbs without limit as the wavelength shrinks, running off the top of the graph. That runaway is the ultraviolet catastrophe.

An object at any temperature emits electromagnetic radiation, and the character of that radiation depends only on the temperature, not on what the object is made of. The idealized emitter is a **blackbody**: a perfect absorber and therefore a perfect emitter, well approximated by a small hole into a heated cavity. Measure the intensity of a blackbody's radiation at each wavelength and you get a smooth curve that peaks at some wavelength and falls off to either side (Figure 39.1). As the temperature rises, the whole curve grows and its peak slides toward shorter wavelengths: the glow shifts from red toward blue. (The total power radiated, the area under the curve, grows as T^4: the Stefan–Boltzmann law of Chapter 37.) The shift of the peak is captured by **Wien's displacement law**:

Key Equation: Wien's displacement law

$$\lambda_{\max} T = 2.898 \times 10^{-3}\,\mathrm{m \cdot K}. \tag{39.1}$$

The trouble was explaining the *shape* of the curve. Applying classical thermodynamics to the cavity's standing electromagnetic waves gave the **Rayleigh–Jeans law**, which agreed with experiment at long wavelengths but diverged catastrophically at short ones: it predicted the intensity should rise without limit as $\lambda \to 0$ (the dashed curve in Figure 39.1). Integrated over all wavelengths, it gave infinite radiated power. This impossible result (the **ultraviolet catastrophe**) meant classical physics simply could not describe a glowing object.

Example 39.1: Peak wavelengths by temperature

Find the peak emission wavelength of (a) human skin at 35°C, (b) a tungsten filament at $2000\ \text{K}$, and (c) the Sun's surface at $5800\ \text{K}$.

Solution

Apply Wien's law, $\lambda_{\max} = (2.898 \times 10^{-3}\ \text{m} \cdot \text{K})/T$.

(a) Skin: $T = 35 + 273 = 308\ \text{K}$, so $\lambda_{\max} = (2.898 \times 10^{-3})/308 = \boxed{9.4\,\mu\text{m}}$, deep infrared, which is why night-vision cameras see body heat.

(b) Tungsten: $\lambda_{\max} = (2.898 \times 10^{-3})/2000 = \boxed{1.4\,\mu\text{m}}$, near infrared, so an incandescent bulb wastes most of its energy as heat, not visible light.

(c) Sun: $\lambda_{\max} = (2.898 \times 10^{-3})/5800 = \boxed{500\,\text{nm}}$, in the middle of the visible band, green-yellow, right where our eyes are most sensitive.

Planck's Radiation Law

Planck found the curve, but only by making an assumption he himself disliked. He proposed that the atomic oscillators in the cavity walls cannot have just any energy; their energy is **quantized**, restricted to integer multiples of a basic quantum proportional to the oscillation frequency:

$$E_n = nhf, \qquad n = 0, 1, 2, \ldots \tag{39.2}$$

where h is a new constant of nature, **Planck's constant**,

$$h = 6.626 \times 10^{-34}\ \text{J} \cdot \text{s}. \tag{39.1.1}$$

With this single restriction, an oscillator can gain energy only one step hf at a time, while the thermal energy available to each oscillator is roughly $k_B T$ (Boltzmann's constant times the temperature). A low-frequency oscillator has small steps, so it is easily excited and shares in the thermal energy just as classical physics assumed. A high-frequency (short-wavelength) oscillator's very first step hf is far larger than $k_B T$, so it is almost never excited (Figure 39.2), and its contribution is suppressed instead of diverging. The catastrophe disappears, and the resulting **Planck radiation law** matches the measured curve at every wavelength and temperature:

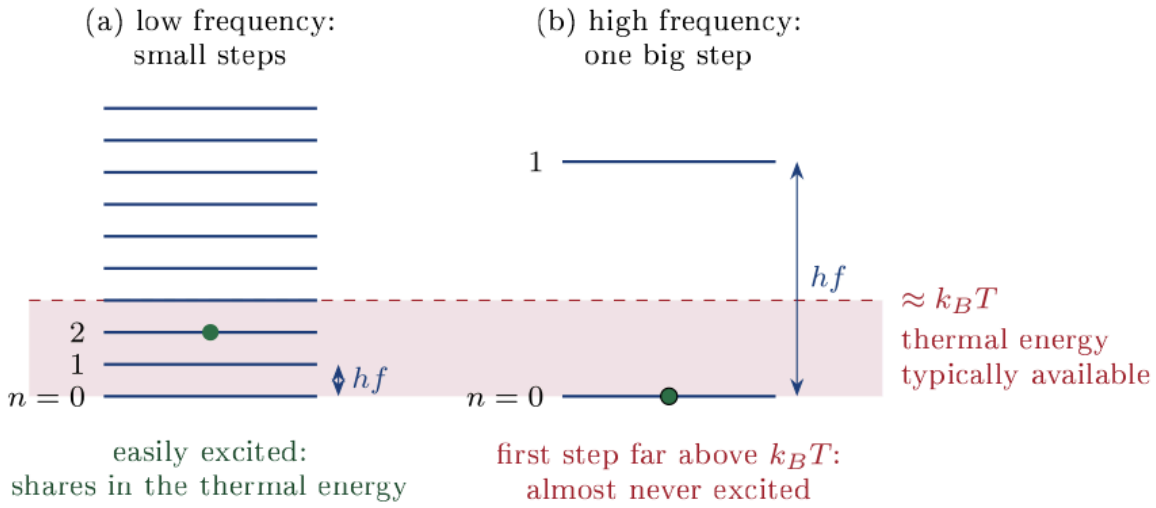


Figure 39.2. Why quantization tames the ultraviolet catastrophe. Each oscillator's energy can change only in steps of hf, and the thermal energy typically available is about k_BT. (a) A low-frequency oscillator has small steps and is easily excited, just as classical physics assumed. (b) A high-frequency oscillator's first step is far above k_BT, so it is almost never excited and radiates almost nothing. That is why the Planck curves turn over at short wavelength.

•Figure description

Why quantization removes the ultraviolet catastrophe, shown as two energy ladders. A pale red band across the bottom of both ladders, topped by a dashed line labelled approximately k sub B T, marks the thermal energy typically available. Panel a, low frequency: a ladder of closely spaced blue levels, n equals 0, 1, 2 and so on, one small step h f apart; several levels lie inside the thermal band, and a green dot sits on level 2. The caption beneath reads easily excited, shares in the thermal energy. Panel b, high frequency: only two levels, n equals 0 and n equals 1, separated by one large step h f that rises far above the thermal band; the green dot sits on the bottom level. The caption beneath reads first step far above k sub B T, almost never excited.

Key Equation: Planck's radiation law

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5 (e^{hc/\lambda k_B T} - 1)}, \tag{39.3}$$

where $k_B = 1.38 \times 10^{-23}\ \text{J/K}$ is the Boltzmann constant.

Planck regarded quantization as a mathematical trick. It fell to Einstein, in 1905, to take it literally and propose that the electromagnetic field itself is quantized: the idea of the photon. Why does the quantum nature of energy never show up in everyday machines? Because h is so tiny that for macroscopic systems the quantum number n is astronomically large, and the steps between levels are far too fine to notice.

Example 39.2: A quantized spring

A $2.00\ \text{kg}$ block on a spring of force constant $k = 25.0\ \text{N/m}$ is pulled $0.400\ \text{m}$ from equilibrium and released. (a) Classically, find the total energy and oscillation frequency. (b) If the energy is quantized as $E = nhf$, find the quantum number n.

Solution

(a) The energy is $E = \frac{1}{2}kA^2 = \frac{1}{2}(25.0)(0.400)^2 = 2.00\ \text{J}$, and the frequency is

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi}\sqrt{\frac{25.0}{2.00}} = \boxed{0.563\,\text{Hz}}. \tag{39.1.2}$$

(b) Treating the oscillator as quantized, $n = E/(hf)$:

$$n = \frac{2.00\,\text{J}}{(6.626 \times 10^{-34}\,\text{J}\cdot\text{s})(0.563\,\text{Hz})} = \boxed{5.4 \times 10^{33}}. \tag{39.1.3}$$

The quantum number is enormous, and a single quantum $hf \approx 4 \times 10^{-34}$ J is utterly imperceptible. This is the *correspondence principle*: for large systems, quantum steps are so fine that the energy looks continuous and classical physics re-emerges.

This page titled 39.1: Blackbody Radiation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.2: Photoelectric Effect

Shine light on a clean metal surface and, under the right conditions, electrons fly off. This **photoelectric effect** baffled classical physics on every count. Classically, light is a wave whose energy depends on its *intensity*, so a bright enough beam of any color should eventually shake electrons loose, and a dim beam should work given time. Experiment said otherwise: below a certain *cutoff frequency* no electrons emerge no matter how intense or how long the light shines; above it, electrons emerge instantly even for feeble light, and their maximum kinetic energy depends on the light's *frequency*, not its intensity.

Einstein explained it all in 1905 with the photon. If light arrives as quanta of energy hf, then one photon ejects one electron, handing over its whole energy. Part goes to freeing the electron from the metal (the **work function** ϕ) and the rest becomes kinetic energy (Figure 39.3):

Key Equation: Photoelectric equation

$$K_{\max} = hf - \phi. \tag{39.4}$$

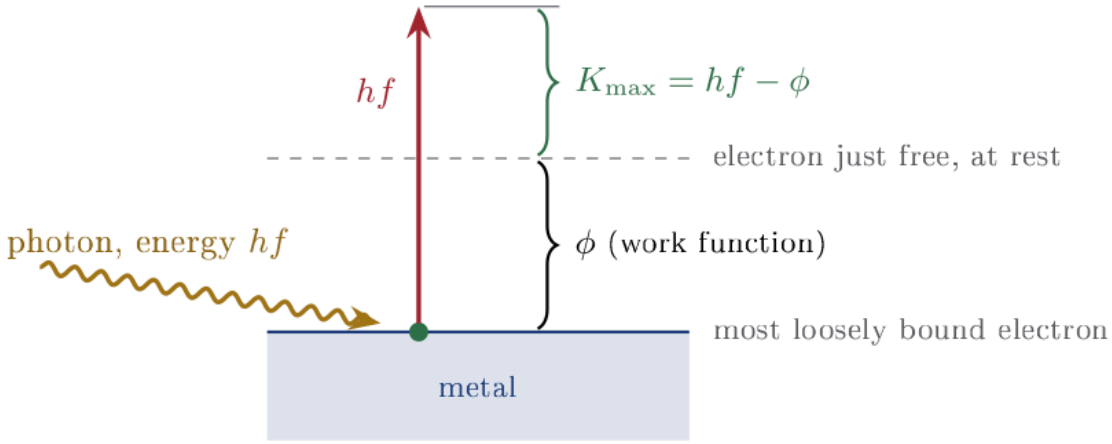


Figure 39.3. Energy bookkeeping in the photoelectric effect. One photon gives its entire energy hf to one electron. The least tightly bound electrons need the work function ϕ just to escape, and whatever is left over is their kinetic energy, so the fastest electrons have $K_{\max} = hf - \phi$. If $hf < \phi$ the arrow never reaches the dashed line and no electron escapes.

•Figure description

An energy diagram of the photoelectric effect. A shaded block at the bottom represents the metal, its top edge marking the energy of the most loosely bound electron. A dashed line higher up marks the energy of an electron just free of the metal and at rest; the gap between the two is the work function phi. A wavy photon arrives from the left and hands its whole energy h f to one electron, drawn as a tall red arrow rising from the top of the metal. Brackets on the right split the arrow's height into two parts: the lower part, phi, is spent escaping the metal, and the upper part, the remainder, is the electron's maximum kinetic energy K max.

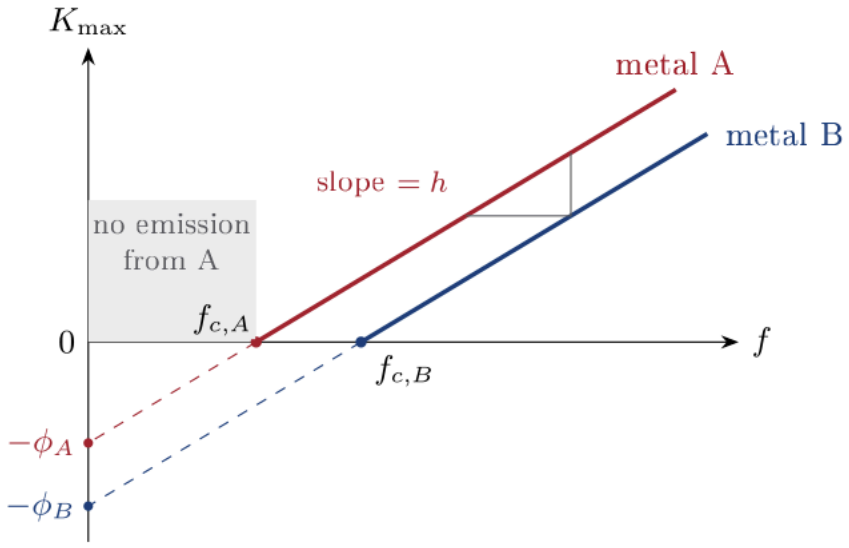


Figure 39.4. Photoelectric effect. The maximum kinetic energy of ejected electrons rises linearly with light frequency, $K_{\max} = hf - \phi$. Below the cutoff frequency $f_c = \phi/h$ no electrons are emitted. Every metal gives a line with the same slope, Planck's constant h; a metal with a larger work function (B) has its line shifted to a higher cutoff frequency.

•Figure description

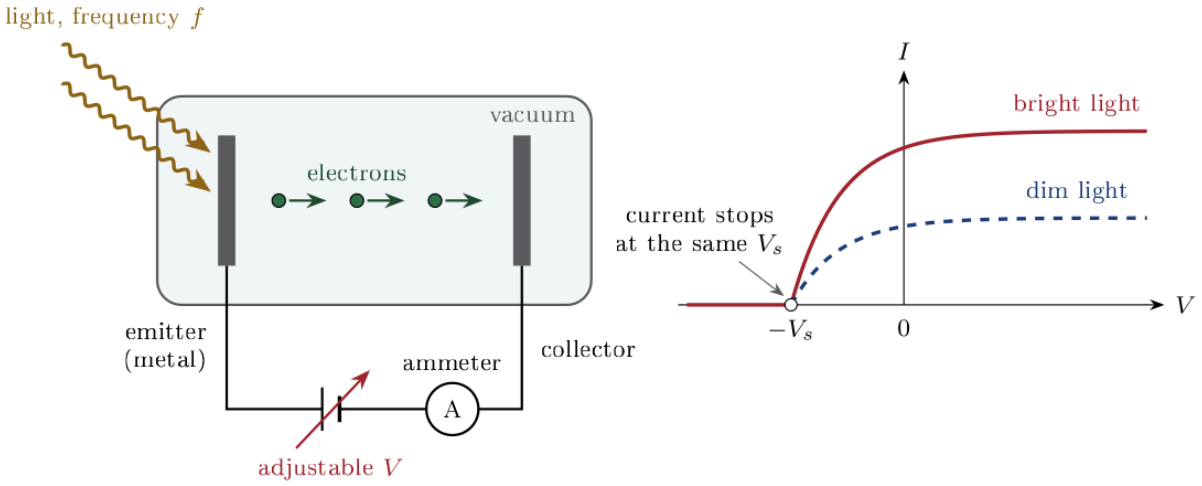
A graph of the maximum kinetic energy of ejected electrons against the frequency of the incident light. Two metals are shown. For metal A the data form a red straight line of slope h, Planck's constant. The line crosses zero at the cutoff frequency f sub c comma A; to the left of that point, shaded as a no-emission region, metal A ejects no electrons at all no matter how bright the light. Extending the line backwards as a dashed segment to the vertical axis gives an intercept of minus phi sub A, the work function of metal A. Metal B, in blue, gives a parallel line with the same slope, starting at a higher cutoff frequency f sub c comma B and extrapolating to a lower intercept, minus phi sub B: a larger work function.

Now everything fits (Figure 39.4). If $hf < \phi$ the photon simply lacks the energy to free an electron, hence a **cutoff frequency** $f_c = \phi/h$, or equivalently a longest **cutoff wavelength** $\lambda_c = hc/\phi$, beyond which nothing happens. Above cutoff, brighter light means *more* photons and thus more electrons, but not more energetic ones; only raising the frequency raises $K_{\max}$. A graph of $K_{\max}$ versus frequency is a straight line of slope h: a way to measure Planck's constant directly. Every metal gives a line with that same slope; only the cutoff frequency and the intercept $-\phi$ change from one metal to another. For these problems the combination $hc = 1240\ \text{eV} \cdot \text{nm}$ is invaluable.

Measuring $K_{\max}$: the stopping potential. How is the electrons' kinetic energy actually measured? In a phototube (Figure 39.5(a)), light falls on a metal emitter inside a vacuum tube, and the ejected electrons cross to a collector, registering as a current I on the ammeter. Making the collector *negative* relative to the emitter pushes the electrons back. As this reverse voltage grows, fewer electrons make it across, until at the **stopping potential** V_s even the fastest ones are turned back and the current drops to zero. At that point the electric field has taken away all of the fastest electron's kinetic energy, so

$$K_{\max} = eV_s. \tag{39.2.1}$$

In electron-volts the bookkeeping is especially simple: a stopping potential of $1.67\ \text{V}$ means $K_{\max} = 1.67\ \text{eV}$. Figure 39.5(b) shows the decisive evidence: brighter light at the same frequency gives a larger current (more electrons) but exactly the same stopping potential (the same $K_{\max}$).



(a) the apparatus (b) current versus voltage, same f

Figure 39.5. Measuring the photoelectric effect. (a) Light ejects electrons from a metal emitter in a vacuum tube; those that reach the collector register as a current. An adjustable voltage can speed them up or, when reversed, turn them back. (b) Current versus voltage at one frequency. Reversing the voltage stops the current at the stopping potential $-V_s$, where $eV_s = K_{\max}$. Brighter light gives more current but the *same* V_s: more electrons, not faster ones.

•Figure description

Measuring the photoelectric effect, in two panels. Panel a, the apparatus: a vacuum tube containing two metal plates, an emitter on the left and a collector on the right. Wavy gold arrows of light of frequency f strike the emitter, and green electrons with arrows cross the tube toward the collector. Outside the tube a wire loop connects the two plates through a battery marked adjustable V, with a red arrow through it, and an ammeter marked A. Panel b, current I against voltage V for light of one frequency: two curves, a solid red one labelled bright light and a dashed blue one labelled dim light. Both are zero for voltages more negative than minus V sub s, rise steeply just above it, and level off at positive voltage, the bright curve leveling off twice as high. A note points to the open circle at minus V sub s: current stops at the same V sub s for both.

Example 39.3: Sodium under ultraviolet light

A sodium surface (work function $\phi = 2.46\ \text{eV}$) is illuminated with light of wavelength $300\ \text{nm}$. (a) Find the maximum kinetic energy of the ejected electrons. (b) Find the cutoff wavelength for sodium.

Solution

(a) The photon energy is $hf = hc/\lambda = (1240\ \text{eV}\cdot\text{nm})/(300\ \text{nm}) = 4.13\ \text{eV}$. Then

$$K_{\max} = hf - \phi = 4.13\,\text{eV} - 2.46\,\text{eV} = \boxed{1.67\,\text{eV}}. \tag{39.2.2}$$

(b) The cutoff wavelength is the longest that still frees an electron, $hf = \phi$:

$$\lambda_c = \frac{hc}{\phi} = \frac{1240\,\text{eV}\cdot\text{nm}}{2.46\,\text{eV}} = \boxed{504\,\text{nm}}. \tag{39.2.3}$$

Light longer than $504\ \text{nm}$ (greenish to red) ejects no electrons from sodium, however bright.

Checkpoint 39.1

Light of a single frequency ejects electrons from a metal. (a) The intensity is doubled at the same frequency. What happens to the number of electrons ejected per second, and to their maximum kinetic energy? (b) The frequency is increased at the same intensity. What happens to each? (c) Which of these results did the classical wave picture of light get wrong?

Answer

Answers to Checkpoints

Checkpoint 39.1. (a) Doubling the intensity doubles the number of photons arriving per second, so twice as many electrons are ejected. Each photon still carries hf, so $K_{\max} = hf - \phi$ is unchanged. (b) $K_{\max}$ increases, since each photon now carries more energy. At the same intensity, more energy per photon means *fewer* photons per second, so somewhat fewer electrons are ejected. (c) The classical picture predicted that brighter light, delivering more energy, would eject electrons with more kinetic energy, and that any frequency would eject electrons if the light were bright enough. Experiment showed that $K_{\max}$ depends on frequency, not intensity, and that below the threshold frequency no electrons emerge at all.

This page titled 39.2: Photoelectric Effect is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.3: Photons

The photoelectric effect forces us to treat light as a stream of particles (**photons**) each carrying a definite quantum of energy set by the frequency,

$$E = hf = \frac{hc}{\lambda}. \tag{39.5}$$

Einstein went further in 1916, arguing that a photon must also carry **linear momentum**. Since a photon is massless, the relativistic relation $E^2 = (pc)^2 + (mc^2)^2$ from Chapter 35 reduces to $E = pc$, so the photon momentum is

Key Equation: Photon energy and momentum

$$E = hf, \qquad p = \frac{E}{c} = \frac{hf}{c} = \frac{h}{\lambda}. \tag{39.6}$$

This is the strange heart of quantum physics: light is *both* a wave and a particle. It diffracts and interferes like a wave (Chapters 33 and 34), yet it delivers its energy and momentum in localized lumps like a particle. Neither picture alone is complete; which face light shows depends on the experiment. The next two sections drive the point home: first by colliding photons with electrons, then by discovering that matter is wavelike too.

This page titled 39.3: Photons is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.4: Compton Shift

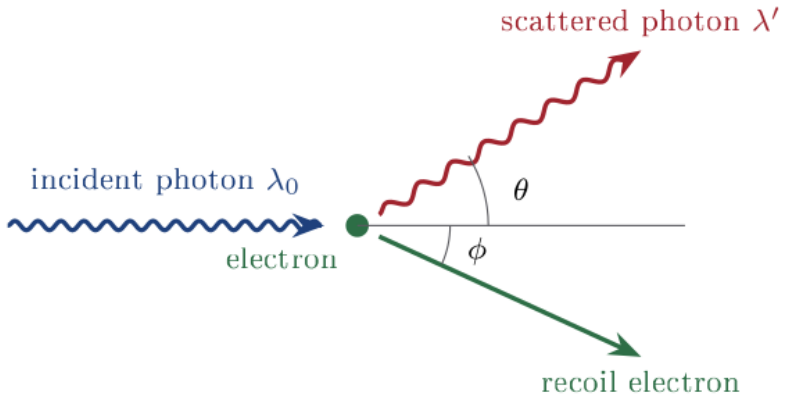


Figure 39.6. Compton scattering. A photon collides with a nearly free electron, scatters at angle θ with longer wavelength λ', and the electron recoils at angle ϕ. The wavelength shift $\Delta\lambda = (h/m_e c)(1 - \cos\theta)$ depends only on the angle: direct proof that photons carry momentum.

•Figure description

Compton scattering, drawn as a collision. A blue wavy arrow, the incident photon of wavelength lambda nought, arrives from the left and strikes a green dot representing an electron at rest. After the collision a red wavy arrow, the scattered photon of longer wavelength lambda prime, drawn with visibly longer wiggles, leaves upward and to the right at an angle theta from the original direction, marked by a small arc. A green arrow shows the electron recoiling downward and to the right at an angle phi below the original direction. The photon has given up momentum to the electron, so it comes away with less energy and a longer wavelength.

If a photon truly carries momentum, then bouncing one off a stationary electron should be an ordinary collision, and the photon, having given up some momentum, should come away with less energy and therefore a *longer* wavelength. In 1923 Arthur Compton fired X-rays at a target and found exactly that: the scattered X-rays had a longer wavelength than the incident beam, by an amount that depended only on the scattering angle (Figure 39.6). Treating it as a relativistic collision that conserves energy and momentum gives the **Compton shift**:

Key Equation: Compton shift

$$\Delta\lambda = \lambda' - \lambda_0 = \frac{h}{m_e c}(1 - \cos\theta), \tag{39.7}$$

where θ is the photon's scattering angle and $\dfrac{h}{m_e c} = 2.43 \times 10^{-3}\ \text{nm}$ is the **Compton wavelength** of the electron.

Notice the shift depends only on the angle, not on the incident wavelength or the target material: a clean signature of photon–electron billiards. It is largest for backscattering ($\theta = 180^\circ$, $\Delta\lambda = 2h/m_e c$) and zero straight ahead ($\theta = 0$). Because the shift is a few thousandths of a nanometer, it is negligible for visible light but plainly measurable for short-wavelength X-rays.

Example 39.4: Scattering X-rays at 45°

X-rays of wavelength $\lambda_0 = 0.200000\ \text{nm}$ scatter from a block of material; the scattered X-rays are observed at 45.0° to the incident beam. Find their wavelength.

Solution

The Compton shift at $\theta = 45.0^\circ$ is

$$\Delta\lambda = \frac{h}{m_e c}(1 - \cos 45.0^\circ) = (2.43 \times 10^{-3}\,\text{nm})(1 - 0.7071) = 7.12 \times 10^{-4}\,\text{nm}. \tag{39.4.1}$$

The scattered wavelength is therefore

$$\lambda' = \lambda_0 + \Delta\lambda = 0.200000\,\text{nm} + 0.000712\,\text{nm} = \boxed{0.200712\,\text{nm}}. \tag{39.4.2}$$

A tiny shift, but Compton measured it precisely, and there is no way to explain it if light is only a wave.

This page titled 39.4: Compton Shift is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.5: de Broglie Wavelength

Here the revolution turns back on itself. If light (long thought a pure wave) behaves like a particle, why should matter (long thought pure particles) not behave like a wave? In 1924 Louis de Broglie made exactly this leap, proposing that any particle of momentum p has an associated wavelength given by the same relation that holds for photons:

Key Equation: de Broglie wavelength

$$\lambda = \frac{h}{p} = \frac{h}{mv}. \tag{39.8}$$

It sounds outrageous, but it is true: beams of electrons diffract and interfere exactly like light when passed through a crystal (the atomic spacing serving as the grating, Chapter 34), and the patterns match $\lambda = h/p$. Electron microscopes exploit this directly, using the tiny wavelength of fast electrons to resolve detail far finer than any light microscope. The reason matter waves escaped notice for so long is, once again, the smallness of h: for anything bigger than an atom the wavelength is preposterously tiny.

Example 39.5: Wavelengths of an electron and a rock

(a) Find the de Broglie wavelength of an electron ($m = 9.11 \times 10^{-31}\ \text{kg}$) moving at $1.00 \times 10^{7}\ \text{m/s}$. (b) Find the de Broglie wavelength of a $50\ \text{g}$ rock thrown at $40\ \text{m/s}$.

Solution

(a) For the electron,

$$\lambda = \frac{h}{mv} = \frac{6.626 \times 10^{-34}}{(9.11 \times 10^{-31})(1.00 \times 10^{7})} = \boxed{7.27 \times 10^{-11}\,\text{m}}, \tag{39.5.1}$$

about $0.073\ \text{nm}$: comparable to atomic spacings, so electrons diffract from crystals.

(b) For the rock,

$$\lambda = \frac{h}{mv} = \frac{6.626 \times 10^{-34}}{(0.050)(40)} = \boxed{3.3 \times 10^{-34}\,\text{m}}. \tag{39.5.2}$$

This is some twenty orders of magnitude smaller than an atomic nucleus: utterly undetectable. That is why a thrown rock shows no wave behavior, even though, in principle, it has a wavelength.

Checkpoint 39.2

An electron and a proton are both moving slowly (not relativistically). (a) If they have the same speed, which has the longer de Broglie wavelength, and by what factor? (b) If they have the same kinetic energy? *(Use $m_p/m_e \approx 1836$ and $p = \sqrt{2mK}$.)*

Answer

Answers to Checkpoints

Checkpoint 39.2. (a) $\lambda = h/mv$, so at equal speeds the lighter particle has the longer wavelength: the electron's is about 1836 times the proton's. (b) With $p = \sqrt{2mK}$ at the same K, $\lambda = h/\sqrt{2mK} \propto 1/\sqrt{m}$. The electron's wavelength is longer by $\sqrt{1836} \approx 43$. Either way, the electron shows its wave nature far more readily.

This page titled 39.5: de Broglie Wavelength is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.6: Heisenberg Uncertainty Principle

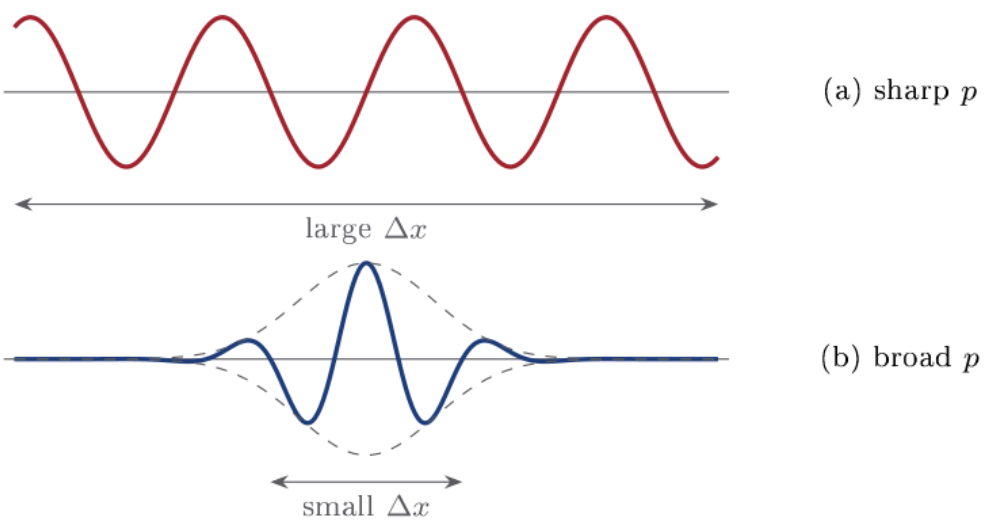


Figure 39.7. The position–momentum tradeoff. (a) A wave of nearly a single wavelength has a sharply defined momentum but extends over a large region (Δx large). (b) Localizing the particle into a small packet (Δx small) requires superposing many wavelengths, so its momentum is broadly spread. You cannot make both sharp at once.

•Figure description

Two waves illustrating the position-momentum tradeoff. Panel (a), labelled sharp p, is a red sine wave of a single wavelength that fills the whole width of the frame at constant amplitude; a double arrow underneath spans the entire frame and is labelled large delta x. Because its wavelength is exactly defined its momentum is sharp, but the particle could be anywhere. Panel (b), labelled broad p, is a blue wave packet: an oscillation confined under a narrow bell-shaped dashed envelope, dying away to nothing on both sides; a much shorter double arrow spans just the packet and is labelled small delta x. Localizing the particle took many superposed wavelengths, so its momentum is now broadly spread.

Wave–particle duality carries a startling price. A perfectly definite momentum corresponds to a pure wave of a single wavelength, which extends through all of space, so a particle with exactly known momentum has a completely undefined position. To localize a particle you must superpose many wavelengths, smearing its momentum (Figure 39.7). You cannot have both sharp at once. Heisenberg made this precise in his **uncertainty principle**: the product of the uncertainties in a particle's position and momentum can never fall below a fixed limit,

Key Equation: Heisenberg uncertainty principle

$$\Delta x\,\Delta p_x \geq \frac{\hbar}{2}, \qquad \Delta E\,\Delta t \geq \frac{\hbar}{2}, \tag{39.9}$$

where $\hbar = h/2\pi = 1.055 \times 10^{-34}\ \mathrm{J\cdot s}$ is the **reduced Planck constant**.

This is not a statement about clumsy instruments; it is a fundamental limit built into nature by the wave character of matter. The better you pin down position, the less you can know about momentum, and vice versa. The energy–time form says the same for energy and the time available to measure it, which is why an atomic state that lasts only a short time cannot have a perfectly sharp energy, giving spectral lines a finite width.

Example 39.6: Confinement: a proton and a dust grain

(a) A proton is confined within a nucleus, a region about $1.0 \times 10^{-14}\ \mathrm{m}$ across. Estimate the minimum uncertainty in its momentum, and the kinetic energy a proton with that much momentum would have, in keV. (b) A $1.0\ \mu\mathrm{g}$ dust grain is located to within $1.0\ \mu\mathrm{m}$ under a microscope. Find the minimum uncertainty in its velocity.

Solution

In each case take Δx as the size of the region, and use $\Delta p_x \geq \hbar/(2\Delta x)$.

(a) The proton.

$$\Delta p_x \geq \frac{\hbar}{2\Delta x} = \frac{1.055 \times 10^{-34}\,\mathrm{J \cdot s}}{2(1.0 \times 10^{-14}\,\mathrm{m})} = \boxed{5.3 \times 10^{-21}\,\mathrm{kg \cdot m/s}}. \tag{39.6.1}$$

A particle whose momentum is uncertain by this much typically has a momentum at least this large, so its kinetic energy is roughly

$$K \approx \frac{(\Delta p_x)^2}{2m_p} = \frac{(5.3 \times 10^{-21}\,\mathrm{kg \cdot m/s})^2}{2(1.67 \times 10^{-27}\,\mathrm{kg})} = 8.3 \times 10^{-15}\,\mathrm{J} \approx \boxed{52\,\mathrm{keV}}. \tag{39.6.2}$$

Confining a particle forces it to move. Squeezing a proton into a nucleus gives it tens of keV at the very least, about ten thousand times the few eV of chemical energies, which is one reason nuclear energies are so large.

(b) The dust grain. With $m = 1.0 \times 10^{-9}$ kg,

$$\Delta v_x = \frac{\Delta p_x}{m} \geq \frac{\hbar}{2m\,\Delta x} = \frac{1.055 \times 10^{-34}\,\mathrm{J \cdot s}}{2(1.0 \times 10^{-9}\,\mathrm{kg})(1.0 \times 10^{-6}\,\mathrm{m})} = \boxed{5.3 \times 10^{-20}\,\mathrm{m/s}}. \tag{39.6.3}$$

At that speed the grain would take about 60 years to drift the width of a single atom. For anything we can see, the uncertainty principle is far below any measurement.

Example 39.7: The natural width of a spectral line

An excited atom radiates with an average lifetime $\tau = 1.0 \times 10^{-8}$ s. Use the uncertainty principle to estimate the resulting spread Δf in the frequency of the emitted light.

Solution

Take the time available as the lifetime, $\Delta t = \tau$. The energy–time relation gives a minimum energy spread $\Delta E \geq \hbar/(2\tau)$. Since a photon's energy is $E = hf$, the energy spread corresponds to a frequency spread $\Delta f = \Delta E/h$:

$$\Delta f \geq \frac{\hbar}{2\tau h} = \frac{1}{4\pi\tau} = \frac{1}{4\pi(1.0 \times 10^{-8}\,\mathrm{s})} = \boxed{8.0 \times 10^{6}\,\mathrm{Hz}}. \tag{39.6.4}$$

This "natural linewidth" of a few megahertz is unavoidable: no spectral line can be infinitely sharp, because no excited state lasts forever. The shorter the lifetime, the broader the line.

Checkpoint 39.3

A classmate claims the uncertainty principle is just a statement that our instruments disturb what they measure: measure position with a photon and you kick the particle's momentum. Explain why this "disturbance" story, while not wrong about disturbance, misses the deeper point: that the limit follows from the wave nature of matter itself and would hold even for an ideal, gentle measurement.

Answer

Answers to Checkpoints

Checkpoint 39.3. The disturbance story treats the particle as having an exact position and momentum that our clumsy measurement spoils. But in quantum mechanics the particle is described by a wave. A wave confined to a small region of space must be built up from a wide range of wavelengths, and since $p = h/\lambda$, that means a wide range of momenta. A wave with a single, sharp wavelength (a definite momentum) extends over all of space. The trade-off $\Delta x\ \Delta p_x \geq \hbar/2$ is a property of the wave itself, which holds before any measurement is made. A perfectly gentle measurement could not beat it, because there is no underlying exact position and momentum to reveal.

This page titled 39.6: Heisenberg Uncertainty Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.S: Introduction to Quantum Mechanics (Summary)

Summary

- **Blackbody radiation:** every body radiates a temperature-dependent spectrum; the peak obeys **Wien's law** $\lambda_{\max}T = 2.898 \times 10^{-3}\ \mathrm{m \cdot K}$. Classical theory (Rayleigh–Jeans) diverged at short λ: the **ultraviolet catastrophe**.
- **Planck's law:** quantizing oscillator energy as $E = nhf$ (with $h = 6.626 \times 10^{-34}\ \mathrm{J \cdot s}$) fits the curve and resolves the catastrophe. $I(\lambda, T) = 2\pi hc^2/[\lambda^5(e^{hc/\lambda k_B T} - 1)]$.
- **Photoelectric effect:** $K_{\max} = hf - \phi$, measured by the stopping potential, $K_{\max} = eV_s$. Below the cutoff $f_c = \phi/h$ (i.e. above $\lambda_c = hc/\phi$) no electrons are ejected; brighter light gives more electrons, not faster ones.
- **Photons:** light comes in quanta of energy $E = hf = hc/\lambda$ and momentum $p = h/\lambda$; light is both wave and particle.
- **Compton shift:** a photon scattering off an electron lengthens by $\Delta\lambda = (h/m_e c)(1 - \cos\theta)$, with Compton wavelength $h/m_e c = 2.43 \times 10^{-3}\ \mathrm{nm}$, proof of photon momentum.
- **de Broglie wavelength:** matter is wavelike too, $\lambda = h/p$; electrons diffract, but macroscopic wavelengths are immeasurably small.
- **Uncertainty principle:** $\Delta x\ \Delta p_x \geq \hbar/2$ and $\Delta E\ \Delta t \geq \hbar/2$ ($\hbar = h/2\pi$), a fundamental limit from the wave nature of matter, not instrument error.

Key Equations

Wien's displacement law

$$\lambda_{\max} T = 2.898 \times 10^{-3}\,\mathrm{m \cdot K} \tag{39.S.1}$$

Planck's quantization of oscillator energy

$$E_n = nhf, \qquad n = 0, 1, 2, \ldots \tag{39.S.2}$$

Planck's radiation law

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5\left(e^{hc/\lambda k_B T} - 1\right)} \tag{39.S.3}$$

Photoelectric equation

$$K_{\max} = hf - \phi \tag{39.S.4}$$

Stopping potential

$$K_{\max} = eV_s \tag{39.S.5}$$

Cutoff frequency and wavelength

$$f_c = \frac{\phi}{h}, \qquad \lambda_c = \frac{hc}{\phi} \tag{39.S.6}$$

Photon energy and momentum

$$E = hf = \frac{hc}{\lambda}, \qquad p = \frac{E}{c} = \frac{h}{\lambda} \tag{39.S.7}$$

Compton shift

$$\Delta\lambda = \lambda' - \lambda_0 = \frac{h}{m_e c}(1 - \cos\theta), \qquad \frac{h}{m_e c} = 2.43 \times 10^{-3}\,\text{nm} \tag{39.S.8}$$

de Broglie wavelength

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{39.S.9}$$

Heisenberg uncertainty principle

$$\Delta x\,\Delta p_x \geq \frac{\hbar}{2}, \qquad \Delta E\,\Delta t \geq \frac{\hbar}{2}, \qquad \hbar = \frac{h}{2\pi} \tag{39.S.10}$$

Useful constants

$$h = 6.626 \times 10^{-34}\,\text{J}\cdot\text{s}, \qquad hc = 1240\,\text{eV}\cdot\text{nm} \tag{39.S.11}$$

This page titled 39.S: Introduction to Quantum Mechanics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

39.E: Introduction to Quantum Mechanics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $h = 6.626 \times 10^{-34}$ J · s, $hc = 1240$ eV · nm, $\hbar = 1.055 \times 10^{-34}$ J · s, $m_e = 9.11 \times 10^{-31}$ kg, electron Compton wavelength $h/m_e c = 2.43 \times 10^{-3}$ nm, and Wien's constant 2.898×10^{-3} m · K.

Section 39.1 — Blackbody Radiation & Planck's Law

39.1 ★ A star's blackbody spectrum peaks at 290 nm. What is its surface temperature?

39.2 ★★ *[Conceptual.]* Explain what the ultraviolet catastrophe was and how Planck's quantization of oscillator energy resolved it.

39.3 ★★ The Sun's thermal spectrum peaks near 500 nm. Use Wien's law to estimate the temperature of its surface.

Sections 39.2–39.3 — Photoelectric Effect & Photons

39.4 ★ A metal has a work function of 4.5 eV. Find its cutoff wavelength.

39.5 ★★ Light of wavelength 250 nm strikes a metal of work function 3.2 eV. Find the maximum kinetic energy of the ejected electrons.

39.6 ★★ *[Conceptual.]* In the photoelectric effect, increasing the *intensity* of light above the cutoff frequency increases the number of ejected electrons but not their maximum kinetic energy. Explain this in terms of photons.

39.7 ★★★ *[Real-world application: solar.]* A photon of green light ($\lambda = 530$ nm) strikes a silicon solar cell. (a) Find the photon energy in eV. (b) Silicon's band gap is 1.1 eV; explain why this photon can free a charge carrier but a far-infrared photon ($\lambda = 10$ μm) cannot.

39.8 ★★ A metal surface has a work function of 2.3 eV.

a. Find the longest wavelength of light that can eject electrons from it.
b. If 400 nm light is used, find the maximum kinetic energy of the ejected electrons.

39.9 ★★ In a photoelectric experiment, light of wavelength 400 nm shines on a metal, and the photocurrent drops to zero when the collector is 0.80 V negative relative to the emitter. (a) What is the maximum kinetic energy of the ejected electrons, in eV? (b) What is the work function of the metal? (c) What stopping potential would 300 nm light require? (d) If the 400 nm light is made twice as bright, what happens to the stopping potential and to the photocurrent?

39.10 ★★ A red LED emits photons of wavelength 640 nm. Find the energy of one photon in (a) joules and (b) electron-volts.

Section 39.4 — Compton Shift

39.11 ★ X-rays scatter from electrons at 90°. Find the Compton shift in wavelength.

39.12 ★★ X-rays of wavelength 0.0500 nm are scattered at 60°. Find the wavelength of the scattered X-rays.

39.13 ★★★ At what scattering angle is the Compton shift equal to the Compton wavelength of the electron?

39.14 ★★ X-rays of wavelength 0.040 nm Compton-scatter from electrons. Find the wavelength of the photons scattered through (a) 90° and (b) 180°.

Section 39.5 — de Broglie Wavelength

39.15 ★ Find the de Broglie wavelength of a proton ($m = 1.67 \times 10^{-27}$ kg) moving at 1.0×10^{6} m/s.

39.16 ★★ *[Real-world application: microscopy.]* Through what potential difference must an electron be accelerated to have a de Broglie wavelength of 0.10 nm (about an atomic spacing)? *(Use $K = \frac{1}{2}mv^2 = p^2/2m$ and $eV = K$.)*

39.17 ★★★ *[Conceptual.]* A baseball and an electron have the same de Broglie wavelength. Which has the greater momentum, and which the greater speed? Explain.

39.18 ★★ (a) Find the de Broglie wavelength of an electron moving at 2.0×10^{6} m/s. (b) Find the de Broglie wavelength of a 0.050 kg golf ball moving at 40 m/s, and explain why its wave nature is never observed.

Section 39.6 — Heisenberg Uncertainty & Synthesis

39.19 ★★ An electron is confined to a region 0.10 nm across (about the size of an atom). Estimate the minimum uncertainty in its momentum.

39.20 ★★★ An excited nuclear state has a lifetime of 1.0×10^{-12} s. Estimate the minimum uncertainty in its energy, in eV.

39.21 ★★★ *[Conceptual — wave or particle?]* Return to the chapter's through-line. For each of the following, state whether it reveals light (or matter) behaving as a *wave* or as a *particle*, and why: (a) the double-slit interference pattern, (b) the photoelectric effect, (c) the Compton shift, (d) electron diffraction from a crystal. What general principle ties them together?

Challenge Problems

39.22 ★★★ *[Real-world application: astronomy — counting photons.]* A cool red giant radiates approximately as a blackbody whose spectrum peaks at 850 nm. An amateur astronomer observes a sixth-magnitude star (the faintest visible to the unaided eye) whose energy flux at Earth is about 1.0×10^{-11} W/m^2, using a telescope of aperture 0.20 m. Take $hc = 1240$ eV · nm.

a. Find the red giant's surface temperature from Wien's law.
b. Find the energy of a photon at the peak wavelength, in eV and in joules.
c. For the sixth-magnitude star (assume its photons also average 850 nm), find the collecting area of the telescope, the power it gathers, and the rate at which photons arrive.
d. The camera takes exposures of 1.0 ms. How many photons arrive per exposure? Photon arrivals are random, so a count of n photons carries a statistical uncertainty of about $\sqrt{n}$. Find the fractional uncertainty per exposure, and explain why astronomers stack long exposures rather than many short ones.

39.23 ★★★ *[Real-world application: medical imaging — scatter in a gamma camera.]* A nuclear-medicine scan injects technetium-99m, which emits gamma rays of energy 140 keV. Some of these photons Compton-scatter off electrons in the patient's tissue before reaching the detector; because they arrive from the wrong direction, they blur the image. Take $hc = 1240$ eV · nm and the Compton wavelength of the electron as 2.43×10^{-3} nm.

a. Find the wavelength of a 140 keV photon.
b. A photon scatters through $45°$. Find the Compton shift, the scattered photon's wavelength, and its energy in keV. How much energy did the recoiling electron take?
c. Repeat for backscattering, $\theta = 180°$.
d. A gamma camera rejects any photon whose measured energy falls outside a window of $\pm 10\%$ around 140 keV. Which of the two scattered photons above is rejected, and which slips through? Using your answers, explain why scatter rejection gets harder as the photon energy is lowered, and why a detector with better energy resolution produces a sharper image.

39.24 ★★★ *[Real-world application: metrology — why atomic clocks use forbidden transitions.]* The hydrogen $2p$ state decays to the ground state in a mean lifetime of $\tau = 1.6$ ns, emitting a photon at 121.6 nm. Because the state does not live forever, its energy is not perfectly sharp: the energy–time uncertainty relation $\Delta E\ \Delta t \geq \hbar/2$ sets a minimum spread, and hence a minimum spread in the frequency of the emitted light. This is the *natural linewidth*.

a. Taking $\Delta t = \tau$, find the minimum energy spread ΔE of the $2p$ state, in joules and in electron-volts.

b. Convert ΔE to a frequency spread Δf. Find the transition frequency f, and the fractional linewidth $\Delta f / f$.
c. Find the corresponding wavelength spread $\Delta \lambda$, in meters. Comment on how it compares with the size of an atom.
d. An optical-lattice clock uses a transition in strontium whose upper state lives about $150\ \mathrm{s}$. Estimate its fractional linewidth at an optical frequency of $4.3 \times 10^{14}\ \mathrm{Hz}$, and explain in one or two sentences why a long-lived "forbidden" transition is exactly what a clock designer wants, and what practical difficulty it creates.

This page titled 39.E: Introduction to Quantum Mechanics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40: A DEEPER LOOK AT QUANTUM MECHANICS

Hold a recent television or tablet close enough to make out the pixels and you are looking at light made by crystals so small that several thousand of them, laid end to end, would barely span the width of a human hair. These are **quantum dots** (specks of semiconductor only a few nanometers across) and the remarkable thing about them is this: a single chemical recipe can be coaxed to glow pure red, pure green, or any hue in between, with the color set not by what the dots are made of but by *how big they are*. Shrink a dot and its light shifts toward the blue; let it grow and the light reddens. No dye, no filter, no change of ingredient, only a change of size.

How can mere size dictate color? The answer is the central idea of this chapter. An electron trapped inside a quantum dot is, to good approximation, a particle confined to a tiny box, and a confined particle cannot have just any energy. It is allowed only a discrete ladder of energies, and, as we will prove, the rungs of that ladder are spaced more widely when the box is smaller. When the electron drops from one rung to a lower one, it emits a photon whose energy, and therefore whose color, equals the gap between rungs. A small dot has widely spaced rungs and emits a high-energy (blue) photon; a large dot has closely spaced rungs and emits a low-energy (red) one. By the end of this chapter you will be able to start from a dot's width and predict the color of the light it gives off.

Learning Objectives

- Use the time-independent Schrödinger equation, normalize a wave function, and compute expectation values such as $\langle x \rangle$.
- Explain why a trapped particle has only discrete (quantized) energies, by matching its matter wave to standing waves.
- Find the allowed energies E_n and wave functions ψ_n of a particle in a one-dimensional infinite well, and the photon energy emitted in a transition.
- Compute the probability of finding the particle in a given region from $\int |\psi|^2 \, dx$.
- Describe how a finite well differs from an infinite one: penetration of the wave function into the walls and slightly lowered energy levels.
- Estimate the probability of quantum tunneling through a barrier with $T \approx e^{-2CL}$.
- Extend the well model to two and three dimensions and recognize the resulting degeneracies.

We will return to the quantum dot as soon as we have the energy levels of a trapped electron, and in the final problems you will predict a dot's color from its size.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

40.1: Schrödinger's Equation and Wave Functions
40.2: Energies of Trapped Electrons
40.3: Infinite Wells
40.4: Finite Wells
40.5: Two- and Three-Dimensional Wells
40.S: A Deeper Look at Quantum Mechanics (Summary)
40.E: A Deeper Look at Quantum Mechanics (Problems)

This page titled 40: A Deeper Look at Quantum Mechanics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.1: Schrödinger's Equation and Wave Functions

In Chapter 39 we found that matter has a wave nature, with a de Broglie wavelength $\lambda = h/p$. The wave is described by a **wave function** $\psi(x)$, and the equation that governs it (the quantum analog of Newton's second law) is the **time-independent Schrödinger equation**. For a particle of mass m moving in one dimension through a region of potential energy $U(x)$,

Key Equation: Time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + U\psi = E\psi, \tag{40.1}$$

where E is the total energy of the particle and $\hbar = h/2\pi$. In the simplest case of a *free* particle, $U = 0$, and the general solution is a combination of traveling waves,

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \tag{40.2}$$

where $k = 2\pi/\lambda$ is the angular wave number. (Equivalent sine-and-cosine forms are often more convenient, and we will use them shortly.)

The wave function itself is not something you can measure. What carries physical meaning is its square: $|\psi(x)|^2$ is a **probability density**, so that the probability of detecting the particle in a small interval dx about x is $|\psi(x)|^2\ dx$. Because the particle must be found *somewhere*, summing this probability over all space must give certainty. This is the **normalization condition**:

Key Equation: Normalization

$$\int_{-\infty}^{\infty} |\psi(x)|^2\, dx = 1. \tag{40.3}$$

Imposing normalization fixes the otherwise-undetermined amplitude constants such as A and B.

Once ψ is known, the average value of any quantity that depends on position (the **expectation value**) follows by weighting that quantity by the probability density. For position itself, and for any function $f(x)$,

$$\langle x\rangle = \int_{-\infty}^{\infty} \psi^*\, x\, \psi\, dx, \qquad \langle f(x)\rangle = \int_{-\infty}^{\infty} \psi^*\, f(x)\, \psi\, dx, \tag{40.4}$$

where ψ^* is the complex conjugate of ψ. The expectation value is the average you would obtain from many measurements on identically prepared particles; it need not be a value the particle can actually have.

Example 40.1: Normalizing a Gaussian wave packet

A particle has the bell-shaped wave function $\psi(x) = Ae^{-ax^2}$ with $a = 1$. The Gaussian integral $\int_{-\infty}^{\infty} e^{-bx^2}\ dx = \sqrt{\pi/b}$ will be useful.

a. Find the value of A that normalizes the wave function.
b. Find the expectation value $\langle x\rangle$ of the particle's position.

Solution

(a) The function is real, so $|\psi|^2 = A^2e^{-2ax^2}$. Normalization, Eq. (40.3), requires

$$\int_{-\infty}^{\infty} A^2e^{-2ax^2}\, dx = A^2\sqrt{\frac{\pi}{2a}} = 1 \quad\Longrightarrow\quad A = \left(\frac{2a}{\pi}\right)^{1/4}. \tag{40.1.1}$$

With $a = 1$,

$$A = \left(\frac{2}{\pi}\right)^{1/4} \approx \boxed{0.893}. \tag{40.1.2}$$

(b) The expectation value of position is

$$\langle x \rangle = \int_{-\infty}^{\infty} \psi^* x \, \psi \, dx = A^2 \int_{-\infty}^{\infty} x \, e^{-2ax^2} \, dx = \boxed{0}. \tag{40.1.3}$$

The integrand $x \; e^{-2ax^2}$ is an odd function integrated over a symmetric interval, so it vanishes. This is exactly what we should expect: the probability density is symmetric about $x = 0$, so the average position is dead center, at the origin.

Checkpoint 40.1

(a) What is the physical meaning of $|\psi(x)|^2 \; dx$? (b) A normalized wave function is multiplied by 2 everywhere. Is it still normalized? (c) A wave function is negative over part of its range. Does that mean a negative probability there?

Answer

Answers to Checkpoints

Checkpoint 40.1. (a) It is the probability of finding the particle between x and $x + dx$ if you look for it; $|\psi|^2$ is a probability *density*. (b) No. Its total probability becomes $\int |2\psi|^2 \; dx = 4 \int |\psi|^2 \; dx = 4$, but the probability of finding the particle somewhere must be exactly 1. (c) No. Probability comes from $|\psi|^2$, which is never negative. The sign of ψ matters for how waves interfere, but not for the probability at a single point.

This page titled 40.1: Schrödinger's Equation and Wave Functions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.2: Energies of Trapped Electrons

A *free* electron, far from any nucleus, can carry any momentum and so have any wavelength and any energy at all; its energy is continuous. However, confine that electron (trap it between walls), and something dramatic happens: its energy becomes **quantized**, restricted to a discrete set of allowed values.

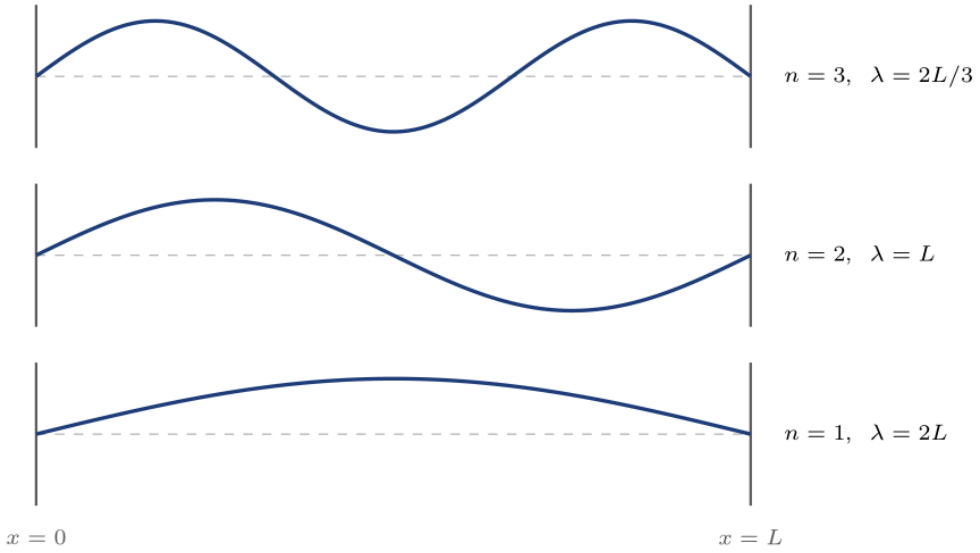


Figure 40.1. The three lowest standing-wave modes ψ_n of a particle trapped between walls a distance L apart. Each mode has a node at both walls, allowing only the wavelengths $\lambda = 2L/n$. These discrete wavelengths are what force the electron's energy to be quantized.

•Figure description

Three stacked diagrams of a particle's matter wave trapped between two walls a distance L apart, drawn as vertical gray lines at x equals 0 and x equals L. The lowest, labeled n equals 1 with wavelength 2 L, is a single blue arch spanning the full width, zero at both walls and peaking in the middle. Above it, n equals 2 with wavelength L shows one full cycle: an upward hump then a downward one, with an extra node at the center. At the top, n equals 3 with wavelength 2 L over 3 shows one and a half cycles and two interior nodes. Every mode is pinned to zero at both walls, which is what restricts the allowed wavelengths.

The reason is the same one that fixes the pitches of a guitar string. A wave confined between two fixed ends can only persist if it forms a **standing wave**, with a node at each wall. For a region of width L, this fits only those wavelengths satisfying

$$L = n\,\frac{\lambda}{2} \qquad \Longrightarrow \qquad \lambda = \frac{2L}{n}, \qquad n = 1, 2, 3, \ldots \tag{40.5}$$

Just as the string sounds only its fundamental and overtones, the trapped electron's matter wave is allowed only the wavelengths in Eq. (40.5) (Figure 40.1), and through $\lambda = h/p$ this restricts its momentum, hence its energy. The integer n is the **quantum number** of the state.

Checkpoint 40.2

A free electron and a trapped electron are otherwise identical. One of them can have an energy of exactly zero; the other cannot have zero energy no matter how it is prepared. Which is which, and why? (Think about what $n = 0$ would do to the standing wave of Eq. (40.5).)

Answer

Answers to Checkpoints

Checkpoint 40.2. The free electron can have zero energy: an unconfined electron at rest has $E = 0$. The trapped electron cannot.

Its wave function must be a standing wave that fits the well, and $n = 0$ would make that wave zero everywhere: no electron at all. The lowest allowed state is $n = 1$, with a nonzero “zero-point” energy. Confinement forces the electron to have some motion.

This page titled 40.2: Energies of Trapped Electrons is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.3: Infinite Wells

The cleanest model of a trapped particle is the **one-dimensional infinite potential well**: a region of width L in which the potential energy is zero inside ($0 < x < L$) but rises to infinity at the walls, so the particle can never escape. With infinitely high walls the wave function is squeezed exactly to zero at $x = 0$ and $x = L$, giving precisely the standing-wave condition of the previous section. (In well diagrams the potential energy is often labeled V rather than U; they mean the same thing.)

Allowed Energies

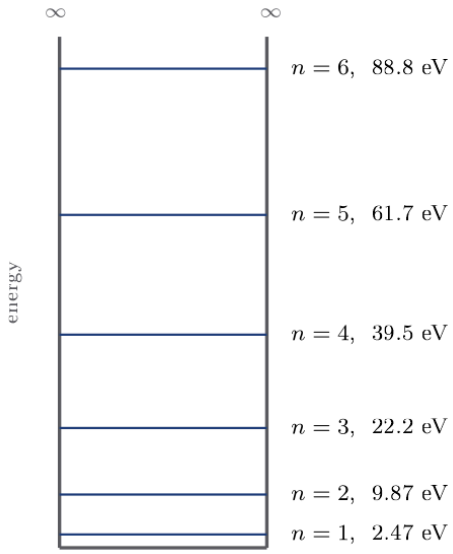


Figure 40.2. Energy-level ladder for an electron in an infinite well (here of width about 0.39 nm, giving $E_1 = 2.47$ eV). Because $E_n \propto n^2$, the rungs spread apart as n increases. A photon is emitted or absorbed only when the electron jumps between two rungs.

•Figure description

An energy-level ladder for an electron in an infinite well, drawn as horizontal blue lines spanning the gap between two tall vertical walls each marked with an infinity symbol. Six levels are labeled n equals 1 through 6, at 2.47, 9.87, 22.2, 39.5, 61.7 and 88.8 electron volts. The rungs are not evenly spaced: because the energy grows as n squared, each gap is wider than the one below it, so the ladder stretches out toward the top.

For a particle inside the well, all its energy is kinetic, so $E = p^2/2m$ and $p = \sqrt{2mE}$. Combining $\lambda = h/p = h/\sqrt{2mE}$ with the standing-wave condition $\lambda = 2L/n$ gives $E = h^2/2m\lambda^2 = n^2h^2/8mL^2$:

Key Equation: Energy levels of an infinite well

$$E_n = \left(\frac{h^2}{8mL^2}\right) n^2, \qquad n = 1, 2, 3, \ldots \tag{40.6}$$

Three features deserve emphasis. First, the energies are *discrete*, labeled by the quantum number n. Second, the lowest allowed energy is not zero but $E_1 = h^2/8mL^2$: the **zero-point energy**; a confined particle is never truly at rest, a direct consequence of the uncertainty principle. Third, the levels go as n^2, so they spread apart as n grows (Figure 40.2). Crucially, $E_n \propto 1/L^2$: the narrower the well, the higher and more widely spaced the energy levels, the fact that makes quantum dots tunable.

When an electron drops from a higher level E_{high} to a lower level E_{low}, the energy lost is carried off by a single photon:

$$\Delta E = E_{\text{high}} - E_{\text{low}} = hf = \frac{hc}{\lambda_{\text{photon}}}. \tag{40.7}$$

This is the bridge between the energy ladder and the color we see, and the heart of the quantum-dot story. Because every level

scales as $1/L^2$, a narrower well has a wider gap between its levels and so emits a higher-energy, bluer photon (Figure 40.3).

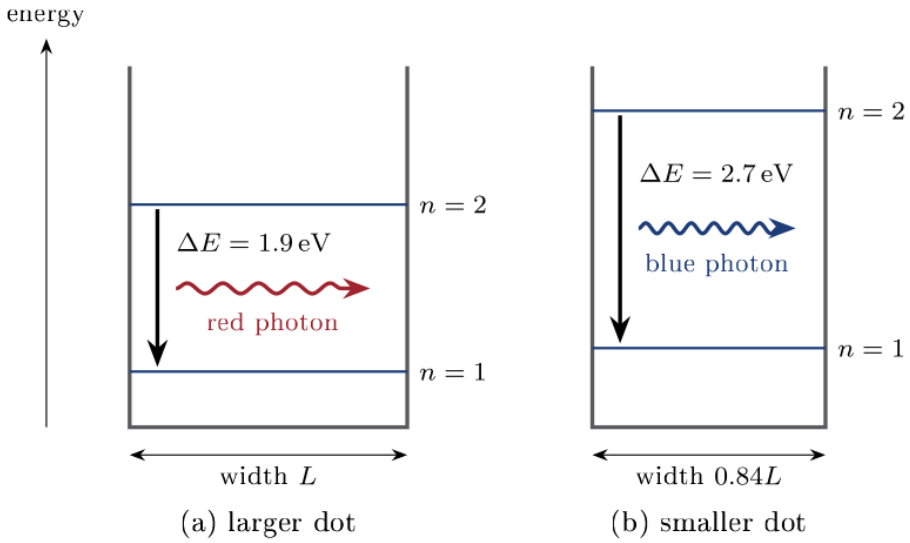


Figure 40.3. Size sets color. Two dots modeled as infinite wells, drawn to the same energy scale. Every level scales as $1/L^2$, so shrinking the width to $0.84L$ raises both levels by a factor of $1/0.84^2 \approx 1.42$ and widens the $n = 2 \to 1$ gap from 1.9 eV (red light) to 2.7 eV (blue light).

•Figure description

Why smaller quantum dots glow bluer, shown as two infinite wells drawn side by side on the same energy scale. Panel a, a larger dot of width L: its n equals 1 and n equals 2 levels are fairly close together. A black arrow points down from n equals 2 to n equals 1, labeled delta E equals 1.9 electron volts, and a red wavy arrow with long wiggles, labeled red photon, leaves the transition. Panel b, a smaller dot of width 0.84 L: both levels sit higher and farther apart. Its downward arrow is labeled delta E equals 2.7 electron volts, and a blue wavy arrow with shorter wiggles is labeled blue photon.

Example 40.2: An electron in a box, and a baseball in a stadium

a. An electron is confined between two impenetrable walls 0.200 nm apart. Find the energies of the states $n = 1, 2, 3$.
b. Find the speed of the electron in the $n = 1$ state.
c. If the electron falls from the second excited state ($n = 3$) to the ground state ($n = 1$), what is the energy of the emitted photon?
d. A 0.500 kg baseball is confined between two rigid walls of a stadium modeled as a box of length 100 m. Find its minimum possible speed.

Solution

(a) With $L = 2.00 \times 10^{-10}$ m, the ground-state energy is

$$E_1 = \frac{h^2}{8m_eL^2} = \frac{(6.63 \times 10^{-34})^2}{8(9.11 \times 10^{-31})(2.00 \times 10^{-10})^2} = 1.51 \times 10^{-18}\,\text{J} = \boxed{9.41\,\text{eV}}. \tag{40.3.1}$$

Since $E_n = n^2E_1$,

$$E_2 = 4E_1 = \boxed{37.6\,\text{eV}}, \qquad E_3 = 9E_1 = \boxed{84.7\,\text{eV}}. \tag{40.3.2}$$

(b) In the well the energy is purely kinetic, $E_1 = \frac{1}{2}m_ev^2$, so

$$v = \sqrt{\frac{2E_1}{m_e}} = \sqrt{\frac{2(1.51 \times 10^{-18})}{9.11 \times 10^{-31}}} = \boxed{1.82 \times 10^6\,\text{m/s}}. \tag{40.3.3}$$

This is about 0.6% of the speed of light: fast, but safely nonrelativistic, so our use of $E = p^2/2m$ was justified.

(c) The photon carries the energy difference,

$$\Delta E = E_3 - E_1 = (9-1)E_1 = 8(9.41\,\text{eV}) = \boxed{75.3\,\text{eV}}. \tag{40.3.4}$$

That corresponds to a wavelength $\lambda = hc/\Delta E = (1240\ \text{eV}\cdot\text{nm})/(75.3\ \text{eV}) \approx 16.5\ \text{nm}$, deep in the ultraviolet: consistent with the wide energy spacing of so small a box.

(d) For the baseball, $m = 0.500\ \text{kg}$ and $L = 100\ \text{m}$, and the minimum energy is the $n = 1$ level:

$$E_1 = \frac{h^2}{8mL^2} = \frac{(6.63\times10^{-34})^2}{8(0.500)(100)^2} = 1.10\times10^{-71}\,\text{J}, \tag{40.3.5}$$

$$v_{\text{min}} = \sqrt{\frac{2E_1}{m}} = \boxed{6.6\times10^{-36}\,\text{m/s}}. \tag{40.3.6}$$

This speed is so absurdly small (the ball would take far longer than the age of the universe to cross a single atomic diameter) that quantization is utterly undetectable for everyday objects. Quantum effects matter only when the box is tiny and the mass is small, exactly the regime of the trapped electron.

The Wave Functions and Probabilities

Solving the Schrödinger equation inside the well (where $U = 0$) with the boundary conditions $\psi(0) = \psi(L) = 0$ selects the sine solutions and quantizes them:

Key Equation: Infinite-well wave functions

$$\psi_n(x) = \sqrt{\frac{2}{L}}\,\sin\!\left(\frac{n\pi}{L}x\right), \qquad n = 1, 2, 3, \ldots \tag{40.8}$$

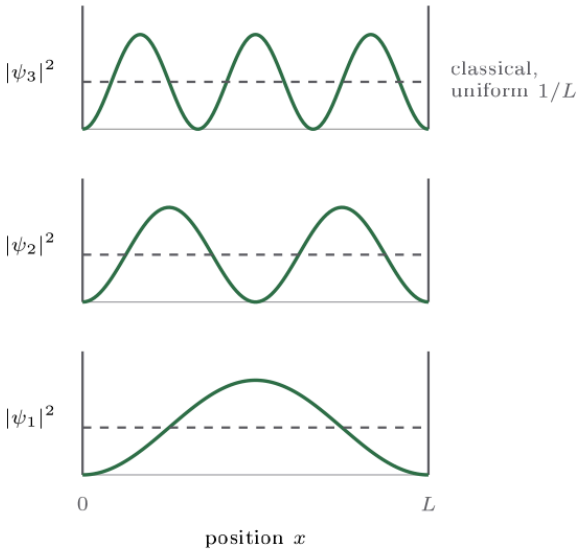


Figure 40.4. Probability densities $|\psi_n|^2$ for the three lowest states of the infinite well. The ground state ($n = 1$) peaks at the center; higher states have n peaks. The area under each curve is 1, as normalization requires. The dashed line is the uniform density $1/L$ of a classical particle bouncing back and forth, which is equally likely to be anywhere.

•Figure description

Three stacked plots of probability density, the square of the wave function, for a particle in an infinite well running from 0 to L. The lowest, labeled psi one squared, is a single green hump peaking at the center of the well and falling to zero at both walls. Above it, psi two squared has two equal humps with a zero exactly at the midpoint, a place the particle is never found. At the top, psi three squared has three humps with two interior zeros. Every curve is non-negative and encloses the same total area of one. A gray dashed horizontal line across each plot, at half the peak height and labeled classical, uniform 1 over L, shows the flat distribution of a classical particle bouncing between the walls.

where the factor $\sqrt{2/L}$ comes from normalizing over $0 \le x \le L$. The probability density is $|\psi_n(x)|^2 = (2/L)\sin^2(n\pi x/L)$ (Figure 40.4). Notice that for $n = 1$ the particle is most likely found near the middle of the well, while for higher n the probability develops several peaks: a pattern utterly unlike the uniform distribution a classical particle bouncing between the walls would have.

To find the probability that the particle lies between x_1 and x_2, integrate the probability density over that interval:

$$P = \int_{x_1}^{x_2} |\psi_n(x)|^2 \, dx = \int_{x_1}^{x_2} \frac{2}{L}\sin^2\!\left(\frac{n\pi}{L}x\right) dx. \tag{40.9}$$

Example 40.3: The average position in the well

A particle of mass m is confined to a one-dimensional infinite well between $x = 0$ and $x = L$. Find the expectation value $\langle x \rangle$ of its position in the state with quantum number n.

Solution

Using Eq. (40.4) with the wave function (40.8),

$$\langle x \rangle = \int_0^L \psi_n^* \, x \, \psi_n \, dx = \frac{2}{L}\int_0^L x \, \sin^2\!\left(\frac{n\pi x}{L}\right) dx. \tag{40.3.7}$$

Writing $\sin^2\theta = \tfrac{1}{2}(1 - \cos 2\theta)$, the cosine term integrates to zero over the full well for any integer n, leaving only

$$\langle x \rangle = \frac{2}{L}\cdot\frac{1}{2}\int_0^L x \, dx = \frac{2}{L}\cdot\frac{1}{2}\cdot\frac{L^2}{2} = \boxed{\frac{L}{2}}. \tag{40.3.8}$$

The result is independent of n: on average the particle sits at the middle of the well. That makes sense, because every $|\psi_n|^2$ in Figure 40.4 is symmetric about the midpoint, so the average position is always the center even though the particle is least likely to be found exactly there for $n > 1$.

Example 40.4: Where is a ground-state electron likely to be?

A ground-state electron ($n = 1$) is trapped in an infinite well of width $L = 100\ \text{pm}$.

a. What is the probability of detecting it in the left one-third of the well?
b. In the middle one-third?
c. In the right one-third?
d. What is the sum of these three probabilities?

Solution

With $n = 1$, Eq. (40.9) gives $P = \frac{2}{L}\int_{x_1}^{x_2} \sin^2(\pi x/L)\ dx$. Carrying out the integral and writing $u = x/L$,

$$P(u_1 \to u_2) = \left[u - \tfrac{1}{2\pi}\sin(2\pi u)\right]_{u_1}^{u_2}. \tag{40.3.9}$$

(a) Left third, $u : 0 \to \tfrac{1}{3}$:

$$P = \tfrac{1}{3} - \frac{1}{2\pi}\sin\frac{2\pi}{3} = 0.3333 - 0.1378 = \boxed{0.196}. \tag{40.3.10}$$

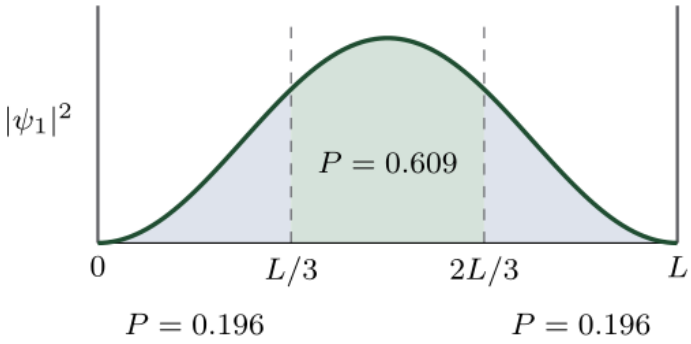
(b) Middle third, $u : \tfrac{1}{3} \to \tfrac{2}{3}$:

$$P = \left(\tfrac{2}{3} - \tfrac{1}{2\pi}\sin\tfrac{4\pi}{3}\right) - \left(\tfrac{1}{3} - \tfrac{1}{2\pi}\sin\tfrac{2\pi}{3}\right) = 0.8045 - 0.1955 = \boxed{0.609}. \tag{40.3.11}$$

(c) Right third, $u : \tfrac{2}{3} \to 1$: by the symmetry of $|\psi_1|^2$ about the center this equals the left third, $\boxed{0.196}$.

(d) The sum is

$$0.196 + 0.609 + 0.196 = \boxed{1.00}. \tag{40.3.12}$$



•Figure description

The ground-state probability density, a single green hump, drawn across an infinite well from 0 to L, with dashed vertical lines at L over 3 and 2 L over 3 dividing it into thirds. The two outer thirds are shaded pale blue and labeled P equals 0.196 each; the middle third, shaded pale green and containing the peak, is labeled P equals 0.609.

The electron is found three times more often in the central third than in either outer third (precisely the shape of the $n = 1$ curve in Figure 40.4) and the three probabilities sum to unity, the check that our wave function is properly normalized.

Checkpoint 40.3

Two electrons are trapped in infinite wells, one of width L and one of width $2L$. Without calculating, decide which electron has the larger ground-state energy and which would emit a higher-frequency (bluer) photon when it drops from $n = 2$ to $n = 1$. Relate your answer to why *smaller* quantum dots glow bluer.

Answer

Answers to Checkpoints

Checkpoint 40.3. The electron in the narrower well, of width L. Infinite-well energies go as $E_n \propto n^2/L^2$, so halving the width multiplies every level by 4: the narrower well has four times the ground-state energy. The $n = 2 \to 1$ photon carries $E_2 - E_1 = 3E_1$, also four times larger, so it has four times the frequency: bluer light. Quantum dots are tiny wells, and smaller dots have more widely spaced levels, so they emit higher-frequency, bluer light.

This page titled 40.3: Infinite Wells is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.4: Finite Wells

Real traps do not have infinitely high walls. A more realistic model is the **finite well**: the potential energy inside is still zero, but outside it rises only to some finite value U (the well depth), with the trapped particle's energy E less than U. We split space into three regions: region I to the left of the well ($x < 0$), region II inside ($0 < x < L$), and region III to the right ($x > L$).

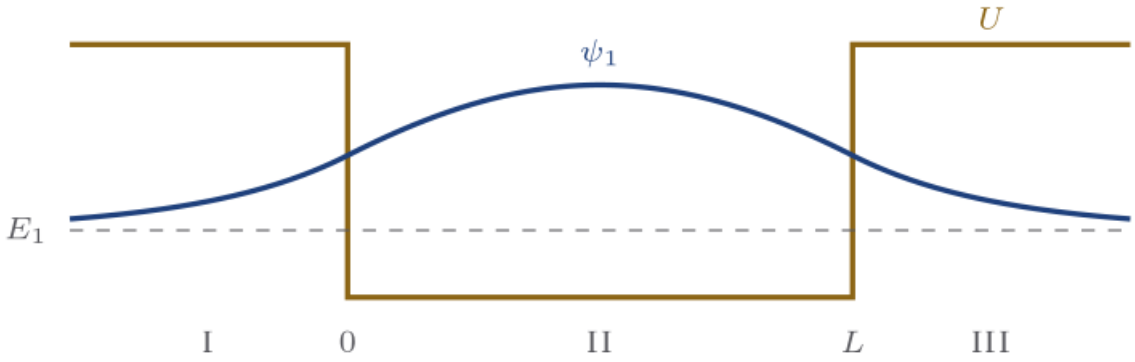


Figure 40.5. Ground-state wave function of a finite well (gold) of depth U. Inside (region II) the wave function is sinusoidal; in the classically forbidden regions I and III it does not vanish but decays exponentially over a depth $\sim 1/C$. The leakage means the particle can be found, with small probability, outside the well.

•Figure description

A finite potential well, drawn in gold: the potential is flat at height U on both sides and drops to zero between x equals 0 and x equals L, dividing space into region I on the left, region II inside, and region III on the right. The blue ground-state wave function is a single smooth arch inside the well, but instead of falling to zero at the walls it continues outward into regions I and III, where it decays away exponentially rather than stopping. That leakage into the classically forbidden regions is the whole point of the figure.

The crucial difference from the infinite well is that the wave function no longer drops abruptly to zero at the walls. It *leaks* into the classically forbidden regions I and III, so there is a small but nonzero chance of finding the particle outside a well it does not have the energy to escape. To see why, rearrange the Schrödinger equation (40.1) as

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = \frac{2m(U-E)}{\hbar^2}\,\psi. \tag{40.10}$$

Inside the well $U = 0 < E$ and the right-hand side is negative, giving oscillating (sinusoidal) solutions just as before. Outside, $U > E$, so the right-hand side is *positive*, and the solutions are exponentials rather than oscillations. Defining

$$C = \frac{\sqrt{2m(U-E)}}{\hbar}, \qquad \text{so that} \qquad \frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = C^2\psi, \tag{40.11}$$

the general solution outside is $\psi = Ae^{Cx} + Be^{-Cx}$. Physical sense (the probability cannot blow up far from the well) forces us to keep only the *decaying* exponential in each outside region:

$$\psi_{\mathrm{I}}(x) = Ae^{Cx} \quad (x < 0), \qquad \psi_{\mathrm{III}}(x) = Be^{-Cx} \quad (x > L), \tag{40.12}$$

each dying away over a characteristic **penetration depth** $1/C$. Inside, region II is sinusoidal, $\psi_{\mathrm{II}}(x) = F\sin kx + G\cos kx$ (Figure 40.5).

We do not solve for all the constants here: that requires matching ψ and its slope at both walls,

$$\begin{aligned} \psi_{\mathrm{I}}(0) &= \psi_{\mathrm{II}}(0), & \psi'_{\mathrm{I}}(0) &= \psi'_{\mathrm{II}}(0), \\ \psi_{\mathrm{II}}(L) &= \psi_{\mathrm{III}}(L), & \psi'_{\mathrm{II}}(L) &= \psi'_{\mathrm{III}}(L), \end{aligned} \tag{40.13}$$

together with normalization, and leads to a transcendental equation best handled numerically. What matters physically is the consequence: because the wave function spreads a little beyond the walls, its effective wavelength is slightly longer than in an

infinite well of the same width, so each energy level sits a little *lower* than the corresponding infinite-well level. A finite well also holds only a finite number of bound states; climb high enough and the particle is no longer trapped at all.

Checkpoint 40.4

In an infinite well the wave function is exactly zero at and beyond the walls; in a finite well it is not. Use this difference to explain, qualitatively, why the finite well's energy levels are lower than the infinite well's for the same width L.

Answer

Quantum Tunneling

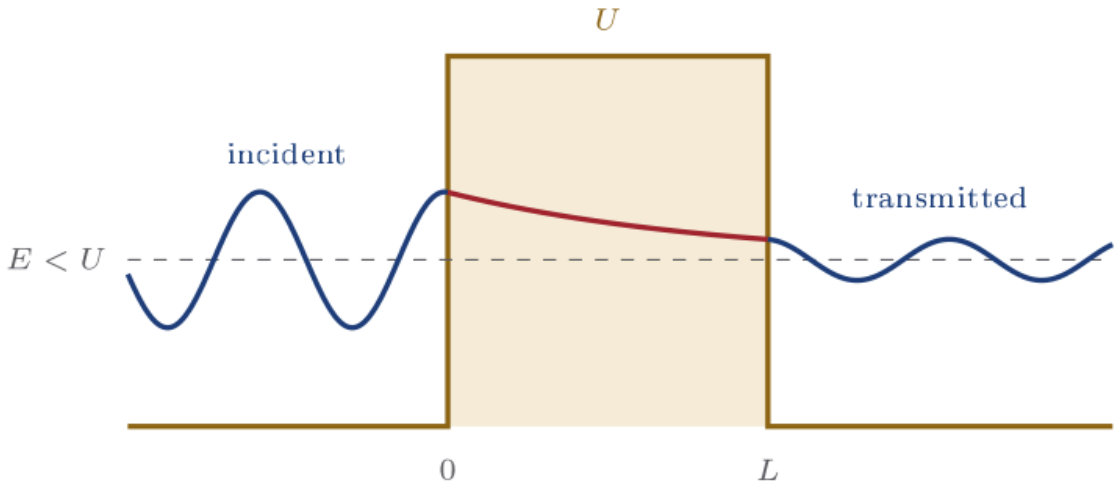


Figure 40.6. Quantum tunneling. A matter wave of energy E strikes a barrier of height $U > E$ and width L. Its amplitude decays exponentially inside the barrier (red) but does not reach zero, so a reduced-amplitude wave emerges on the far side and the particle may be transmitted. Transmission falls off steeply with barrier width.

•Figure description

Quantum tunneling through a barrier. A gold rectangular barrier of height U and width L stands between two flat regions, with a dashed horizontal line marking the particle's energy E, which is below U. On the left a blue wave of constant amplitude travels toward the barrier, labeled incident. Inside the barrier a red curve decays exponentially but does not reach zero before the far wall. On the right the wave resumes oscillating, labeled transmitted, at a visibly smaller amplitude than the incident wave. Some probability gets through a wall the particle has too little energy to climb.

The leakage of the wave function into a forbidden region has a spectacular consequence. Suppose the forbidden region is not infinitely thick but a **barrier** of finite width L: a wall of potential energy U higher than the particle's energy E. The exponentially decaying wave function does not reach zero before the far side of the barrier; a small amplitude survives, and beyond the barrier the wave resumes oscillating. The particle has a real, nonzero chance of appearing on the far side of a wall it classically could never surmount. This is **quantum tunneling** (Figure 40.6).

Classically this is impossible: a ball rolled at a hill too tall always rolls back. Quantum mechanically, the wave function reaches into the forbidden region, so the particle has some chance of being found beyond it, and that chance is appreciable only if the barrier is thin.

We describe the outcome with two probabilities: the **transmission coefficient** T, the probability the particle gets through, and the **reflection coefficient** R, the probability it bounces back. Since it must do one or the other,

$$T + R = 1. \tag{40.14}$$

When the barrier is thick or tall enough that $T \ll 1$, an excellent approximation for the transmission coefficient is

Key Equation: Barrier transmission (tunneling)

$$T \approx e^{-2CL}, \qquad C = \frac{\sqrt{2m(U-E)}}{\hbar}, \tag{40.15}$$

with L the barrier width. The exponential makes T ferociously sensitive to both the width L and the energy shortfall $U - E$: a small change in either can swing the transmission by many orders of magnitude.

Why it matters. This exponential sensitivity is the working principle of the **scanning tunneling microscope** (STM), an instrument that images individual atoms. A fine metal tip is held so close to a surface that electrons tunnel across the vacuum gap; because T depends so steeply on the gap width, even an atom-sized bump changes the tunneling current sharply, and tracking that current maps the surface atom by atom (Figure 40.7). Tunneling also lets alpha particles escape nuclei and lets protons fuse in the Sun's core despite their mutual repulsion: it is one of the most consequential effects in all of physics.

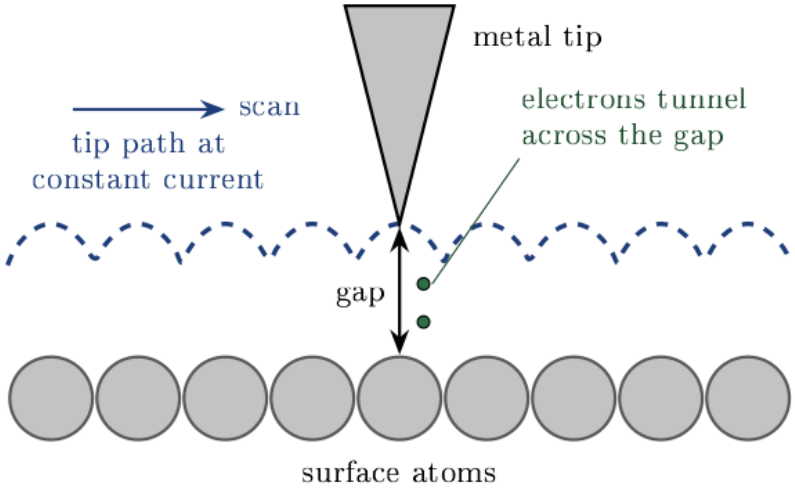


Figure 40.7. A scanning tunneling microscope. Electrons tunnel across the vacuum gap between a sharp tip and the surface. Because $T \approx e^{-2CL}$, the current changes sharply with the gap, so a feedback loop that holds the current constant must raise and lower the tip over each atom. The tip's path (dashed) is a map of the surface.

•Figure description

A scanning tunneling microscope, side view. A row of nine gray circles represents surface atoms. A sharp gray metal tip points straight down at one atom, separated from it by a small vacuum gap marked with a double arrow. Two small green dots in the gap are labeled electrons tunnel across the gap. A blue dashed line runs just above the row of atoms, rising over each atom and dipping between them, and passes through the tip's apex; it is labeled tip path at constant current, with an arrow labeled scan pointing to the right.

Example 40.5: Tunneling through a thin and a thick barrier

A $30\ \text{eV}$ electron is incident on a square barrier of height $40\ \text{eV}$.

a. Find the probability that it tunnels through if the barrier is $1.0\ \text{nm}$ wide.
b. Find the probability if the barrier is only $0.10\ \text{nm}$ wide.

Solution

The energy shortfall is $U - E = 40 - 30 = 10\ \text{eV} = 1.602 \times 10^{-18}\ \text{J}$. The decay constant is

$$C = \frac{\sqrt{2m_e(U-E)}}{\hbar} = \frac{\sqrt{2(9.11\times10^{-31})(1.602\times10^{-18})}}{1.055\times10^{-34}} = 1.62\times10^{10}\,\text{m}^{-1}. \tag{40.4.1}$$

(a) For $L = 1.0 \times 10^{-9}\ \text{m}$, the exponent is $2CL = 2(1.62\times10^{10})(1.0\times10^{-9}) = 32.4$, so

$$T \approx e^{-32.4} = \boxed{8.6 \times 10^{-15}}. \tag{40.4.2}$$

About one electron in 10^{14} gets through: effectively never, for a barrier a nanometer thick.

(b) Shrinking the barrier tenfold to $L = 1.0 \times 10^{-10}\ \mathrm{m}$ cuts the exponent to $2CL = 3.24$:

$$T \approx e^{-3.24} = \boxed{0.039}. \tag{40.4.3}$$

Now nearly 4% of the electrons tunnel through. A tenfold reduction in width raised the transmission by more than twelve orders of magnitude: the very sensitivity that lets a scanning tunneling microscope feel out single atoms.

Checkpoint 40.5

A scanning tunneling microscope's tip drifts very slightly farther from the surface. Using Eq. (40.15), explain qualitatively whether the tunneling current rises or falls, and why even a tiny drift produces a large, easily measured change.

Answer

Answers to Checkpoints

Checkpoint 40.4. In a finite well the wave function does not stop dead at the walls; it leaks a short way into them, decaying exponentially. The electron therefore effectively occupies a region a little wider than L. A wider region fits a standing wave of longer wavelength, and a longer wavelength means a smaller momentum and a lower kinetic energy. So each level sits somewhat below the corresponding level of an infinite well of the same width.

Checkpoint 40.5. It falls. The probability of tunneling through a barrier of thickness L goes as e^{-2CL}. Moving the tip farther away makes the gap (the barrier) thicker, so the current decreases. Because the dependence is *exponential*, with C of order $10^{10}\ \mathrm{m}^{-1}$ for typical barriers, a change in gap of a fraction of an atom's diameter changes the current by a large factor. That sensitivity is what lets an STM map a surface atom by atom.

This page titled 40.4: Finite Wells is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.5: Two- and Three-Dimensional Wells

Nothing forces a trap to be one-dimensional. An electron in a quantum dot or in a tiny conducting island is confined in all directions, and the well model extends naturally. Consider a two-dimensional rectangular well with impenetrable walls and sides L_x and L_y. The standing-wave condition must now be met independently along each axis, so two quantum numbers n_x and n_y appear, and the wave function is a product of sines:

$$\psi_{n_x,n_y}(x,y) = \frac{2}{\sqrt{L_x L_y}} \sin\left(\frac{n_x \pi}{L_x} x\right) \sin\left(\frac{n_y \pi}{L_y} y\right). \tag{40.16}$$

The energy is simply the sum of the energies associated with each direction:

Key Equation: Energy levels of a 2D / 3D box

$$E_{n_x,n_y} = \frac{h^2}{8m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2}\right), \qquad E_{n_x,n_y,n_z} = \frac{h^2}{8m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right), \tag{40.17}$$

with each quantum number running over $1, 2, 3, \ldots$ independently.

Something new emerges when the box is symmetric. In a square box ($L_x = L_y = L$), write $E_0 = h^2/8mL^2$; then $E_{n_x,n_y} = (n_x^2 + n_y^2)E_0$. The ground state $(1,1)$ has energy $2E_0$. But the next level can be reached two ways ($(2,1)$ or $(1,2)$) and *both* give $5E_0$. Two physically distinct states share one energy. This is called **degeneracy**, and it arises directly from the symmetry of the box; deform the box so $L_x \neq L_y$ and the two states split apart in energy. Figure 40.8 shows why: $\psi_{1,2}$ is simply $\psi_{2,1}$ turned through $90°$, and in a square box the turned pattern fits just as well. Degeneracy will reappear, and matter enormously, when we treat the hydrogen atom in Chapter 41.

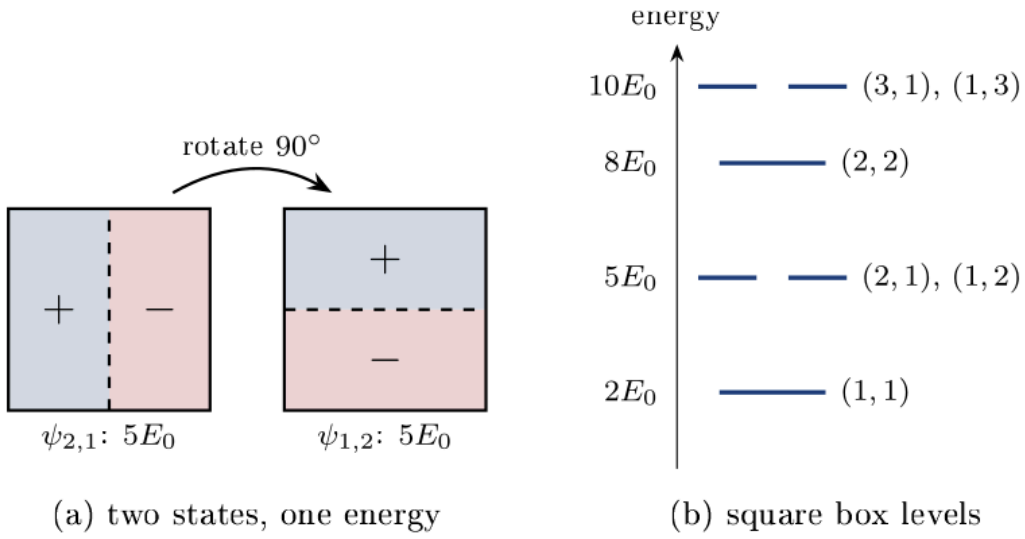


Figure 40.8. Degeneracy in a square box. (a) The states $(2,1)$ and $(1,2)$ have the same pattern turned through $90°$ (dashed lines are nodes, where $\psi = 0$), so in a square box they have the same energy, $5E_0$. (b) The lowest levels of the square box. Pairs of short lines mark doubly degenerate levels.

•Figure description

Degeneracy in a square box, in two panels. Panel a shows two square boxes. In the first, the state psi 2 1 is divided by a vertical dashed nodal line into a left half marked plus, shaded pale blue, and a right half marked minus, shaded pale red. In the second, psi 1 2 is divided by a horizontal dashed nodal line into a top half marked plus and a bottom half marked minus. A curved arrow labeled rotate 90 degrees links them, and both are labeled 5 E nought. Panel b is an energy-level diagram for the square box: a single line at 2 E nought for state 1 1, a pair of short side-by-side lines at 5 E nought for states 2 1 and 1 2, a single line at 8 E nought for state 2 2, and a pair at 10 E nought for states 3 1 and 1 3.

Example 40.6: Energy levels of a square corral

An electron is confined to a square two-dimensional well of side L. In units of $E_0 = h^2/8mL^2$, find the energies of the four lowest levels and state the degeneracy of each.

Solution

The energy is $E_{n_x,n_y} = (n_x^2 + n_y^2)E_0$. Listing the smallest values of $n_x^2 + n_y^2$:

$$\begin{aligned} (1,1): &\ 2E_0 \quad (\text{degeneracy } 1), \\ (2,1),(1,2): &\ 5E_0 \quad (\text{degeneracy } 2), \\ (2,2): &\ 8E_0 \quad (\text{degeneracy } 1), \\ (3,1),(1,3): &\ 10E_0 \quad (\text{degeneracy } 2). \end{aligned} \tag{40.5.1}$$

So the four lowest energies are $\boxed{2E_0,\ 5E_0,\ 8E_0,\ 10E_0}$, with the $5E_0$ and $10E_0$ levels each *doubly degenerate*. Notice that, unlike the one-dimensional well, the spacing here is irregular: a hallmark of multidimensional confinement, and a reminder that the simple n^2 ladder is special to one dimension.

Checkpoint 40.6

An electron is confined to a cubical box of side L. In units of $E_0 = h^2/8mL^2$, what are the energy and degeneracy of (a) the ground state and (b) the first excited level? (c) If one side of the box is stretched slightly, what happens to the degeneracy in (b)?

Answer

Answers to Checkpoints

Checkpoint 40.6. With $L_x = L_y = L_z = L$, $E = (n_x^2 + n_y^2 + n_z^2)E_0$. (a) The ground state is $(1,1,1)$, with $E = 3E_0$, nondegenerate. (b) The first excited level has one quantum number equal to 2: $(2,1,1)$, $(1,2,1)$, and $(1,1,2)$, all with $E = 6E_0$, so it is triply degenerate. (c) The three states no longer share one energy. The degeneracy arose from the box's symmetry; stretching one side breaks the symmetry, and the state with the 2 along the stretched side drops slightly below the other two, which remain degenerate with each other.

This page titled 40.5: Two- and Three-Dimensional Wells is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.S: A Deeper Look at Quantum Mechanics (Summary)

Summary

- The **time-independent Schrödinger equation** $-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+U\psi=E\psi$ governs the wave function ψ. Only $|\psi|^2$ is physical: it is a probability density, normalized by $\int|\psi|^2\ dx=1$, and averages follow from $\langle f(x)\rangle=\int\psi^* f(x)\psi\ dx$.
- A **trapped** particle has *quantized* energies because its matter wave must form a standing wave, $\lambda=2L/n$. A free particle's energy is continuous.
- **Infinite well**: $E_n=(h^2/8mL^2)n^2$ and $\psi_n=\sqrt{2/L}\sin(n\pi x/L)$. The lowest energy E_1 is nonzero (zero-point energy); levels go as n^2 and as $1/L^2$ (smaller box ⇒ higher, wider-spaced levels). A transition emits a photon with $\Delta E=E_{\text{high}}-E_{\text{low}}=hf$. Probabilities come from $\int_{x_1}^{x_2}|\psi_n|^2\ dx$; $\langle x\rangle=L/2$ for every n.
- **Finite well**: the wave function decays exponentially into the walls ($\psi\sim e^{\pm Cx}$, $C=\sqrt{2m(U-E)}/\hbar$) rather than vanishing, so the particle can be found just outside. Energy levels are slightly lower than the infinite well's, and only finitely many bound states exist.
- **Quantum tunneling**: a particle can pass through a barrier with $E<U$. For $T\ll 1$, $T\approx e^{-2CL}$ with $C=\sqrt{2m(U-E)}/\hbar$, and $T+R=1$. Transmission is extraordinarily sensitive to barrier width: the basis of the scanning tunneling microscope.
- **Two and three dimensions**: one quantum number per axis, $E=\frac{h^2}{8m}\sum_i n_i^2/L_i^2$. Symmetric boxes produce **degeneracy**: distinct states with equal energy.

Key Equations

Time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+U\psi=E\psi \tag{40.S.1}$$

Normalization and expectation values

$$\int_{-\infty}^{\infty}|\psi(x)|^2\,dx=1$$
$$\langle f(x)\rangle=\int_{-\infty}^{\infty}\psi^*\,f(x)\,\psi\,dx \tag{40.S.2}$$

Standing-wave condition for a trapped particle

$$\lambda=\frac{2L}{n},\qquad n=1,2,3,\ldots \tag{40.S.3}$$

Energy levels of an infinite well

$$E_n=\left(\frac{h^2}{8mL^2}\right)n^2\qquad(E_n\propto n^2,\ E_n\propto 1/L^2) \tag{40.S.4}$$

Infinite-well wave functions

$$\psi_n(x)=\sqrt{\frac{2}{L}}\sin\!\left(\frac{n\pi}{L}x\right),\qquad\langle x\rangle=\frac{L}{2}\text{ for every }n \tag{40.S.5}$$

Photon emitted in a transition

$$\Delta E = E_{\text{high}} - E_{\text{low}} = hf = \frac{hc}{\lambda_{\text{photon}}} \tag{40.S.6}$$

Probability of finding the particle in an interval

$$P = \int_{x_1}^{x_2} |\psi_n(x)|^2 \, dx \tag{40.S.7}$$

Decay constant outside a well or barrier

$$C = \frac{\sqrt{2m(U-E)}}{\hbar}, \qquad \text{penetration depth } = \frac{1}{C} \tag{40.S.8}$$

Barrier transmission (tunneling)

$$T \approx e^{-2CL}, \qquad T + R = 1 \tag{40.S.9}$$

Energy levels of a 2D / 3D box

$$E_{n_x,n_y,n_z} = \frac{h^2}{8m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right) \tag{40.S.10}$$

This page titled 40.S: A Deeper Look at Quantum Mechanics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

40.E: A Deeper Look at Quantum Mechanics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $h = 6.63 \times 10^{-34}$ J·s, $\hbar = 1.055 \times 10^{-34}$ J·s, $m_e = 9.11 \times 10^{-31}$ kg, $c = 3.00 \times 10^{8}$ m/s, $1\ \text{eV} = 1.602 \times 10^{-19}$ J, and the handy combination $hc = 1240\ \text{eV}\cdot\text{nm}$.

Section 40.1 — Schrödinger's Equation and Wave Functions

40.1 ★ *[Conceptual.]* Explain in your own words why the wave function ψ itself is not measurable, while $|\psi|^2$ is, and state what physical quantity $|\psi|^2$ represents.

40.2 ★★ A particle's wave function is $\psi(x) = Ae^{-ax^2}$ with $a = 2.0\ \text{nm}^{-2}$. Using

$$\int_{-\infty}^{\infty} e^{-bx^2}\,dx = \sqrt{\pi/b}, \tag{40.E.1}$$

find the normalization constant A.

40.3 ★★ For the wave function of the previous problem, argue from symmetry that $\langle x\rangle = 0$, then state whether $\langle x^2\rangle$ is zero or positive and why.

40.4 ★★ A particle is confined to the region $0 \le x \le a$, where its wave function is $\psi(x) = Ax(a - x)$; outside that region $\psi = 0$.

a. Find the normalization constant A in terms of a.
b. Find $\langle x\rangle$, and explain how you could have predicted the result without integrating.
c. Find the probability that the particle is in the left quarter of the region, $0 < x < a/4$.
d. This parabola is a common stand-in for the ground state of an infinite well of width a (Section 40.3), for which the same probability is 0.091. By what percentage does the parabola's estimate differ?

Sections 40.2–40.3 — Trapped Electrons and Infinite Wells

40.5 ★ An electron is trapped in a one-dimensional infinite well of width $0.300\ \text{nm}$. Find the energy (in eV) of the ground state.

40.6 ★★ For the well of the previous problem, find the energies of the $n = 2$ and $n = 3$ states and the wavelength of the photon emitted in the $n = 3 \to n = 1$ transition.

40.7 ★★ *[Real-world application: engineering.]* An electron in an infinite well emits a photon of wavelength $45\ \text{nm}$ when it drops from $n = 2$ to $n = 1$. Find the width L of the well.

40.8 ★★ A ground-state electron is in an infinite well of width L. Find the probability of detecting it in the central half of the well ($L/4 < x < 3L/4$).

40.9 ★★ *[Conceptual.]* Sketch $|\psi_2|^2$ for the infinite well and mark the points where the electron is *never* found. Explain how a classical particle's probability distribution would differ.

40.10 ★★★ Show, by evaluating $\int_0^L |\psi_n|^2\,dx$ with $\psi_n = \sqrt{2/L}\sin(n\pi x/L)$, that the wave function (40.8) is correctly normalized for any positive integer n.

40.11 ★★ An electron is confined to a one-dimensional infinite well $0.50\ \text{nm}$ wide. Find the energies (in eV) of (a) the ground state and (b) the first excited state.

40.12 ★★ *[Real-world application: spectroscopy.]* An electron is trapped in a one-dimensional infinite well $1.00\ \text{nm}$ wide.

a. Find the ground-state energy E_1 in eV.
b. Considering only the levels $n = 1$ through 4, list all six possible downward transitions and the wavelength of the photon each

one emits.
c. Which of these photons are visible ($400\ \text{nm}$ to $700\ \text{nm}$)?
d. If the well were twice as wide, by what factor would every one of these wavelengths change?

Section 40.4 — Finite Wells

40.13 ★ *[Conceptual.]* List two physical ways a finite well differs from an infinite well of the same width, and explain each in one sentence.

40.14 ★★ An electron is bound in a finite well of depth $U = 40\ \text{eV}$ in a state with energy $E = 25\ \text{eV}$. Find the decay constant C outside the well and the penetration depth $1/C$.

40.15 ★★ *[Conceptual.]* Compare a finite square well with an infinite one. (a) How do the bound-state wave functions differ in the region outside the well? (b) Why does a finite well support only a finite number of bound states?

Section 40.4.1 — Quantum Tunneling

40.16 ★★ A $5.0\ \text{eV}$ electron is incident on a $10.0\ \text{eV}$ barrier of width $0.20\ \text{nm}$. Estimate the transmission coefficient T.

40.17 ★★★ *[Real-world application: instrumentation.]* In a scanning tunneling microscope an electron tunnels across a vacuum gap with $U - E = 4.0\ \text{eV}$.

a. Find the decay constant C.
b. By what factor does the tunneling current ($\propto T$) change if the gap widens by just $0.10\ \text{nm}$?
c. Explain how this sensitivity lets the instrument resolve individual atoms.

Section 40.5 — Two- and Three-Dimensional Wells & the Quantum-Dot Callback

40.18 ★★ An electron is confined to a square two-dimensional well of side $L = 0.50\ \text{nm}$. Find the energy of the ground state $(1, 1)$ in eV, and identify the degeneracy of the first excited level.

40.19 ★★★ A cubical box of side L holds an electron. In units of $E_0 = h^2/8mL^2$, find the energies and degeneracies of the three lowest levels, taking care to count all (n_x, n_y, n_z) combinations.

40.20 ★★★ *[Real-world application: nanotechnology — the quantum-dot callback.]* Reconsider the quantum dot from the chapter opener, modeled as an electron in a one-dimensional infinite well whose width L equals the dot's diameter.

a. A red-emitting dot releases a photon of energy $1.9\ \text{eV}$ when an electron drops from $n = 2$ to $n = 1$. Show that the required well width is $L = \sqrt{3h^2/(8m_e\,\Delta E)}$, and evaluate it.
b. A second dot is made half as wide. Find its $n = 2 \to n = 1$ photon energy and state whether its light is redder or bluer than the first dot's.
c. Explain, using $E_n \propto 1/L^2$, why manufacturers tune a display's colors purely by controlling dot size, and confirm this matches the opener's claim that smaller dots glow bluer.

Challenge Problems

40.21 ★★★ *[Real-world application: electronics — how long flash memory remembers.]* A flash memory cell stores a bit as a packet of electrons trapped on a "floating gate" surrounded by silicon dioxide. The electrons are not held there by a wall they cannot climb: they are held by a barrier they can only *tunnel* through. Model the oxide as a rectangular barrier with $U - E = 3.1\ \text{eV}$, and take the trapped electron to strike the barrier $\nu = 1 \times 10^{15}$ times per second, escaping with probability $T \approx e^{-2CL}$ on each attempt.

a. Find the decay constant C inside the barrier.
b. For an oxide thickness $L = 3.0\ \text{nm}$, find T and the mean retention time $t = 1/(\nu T)$. Express the retention time in years.

c. Now shrink the oxide to $L = 2.5\ \text{nm}$: a reduction of about two atomic layers. Recompute T and the retention time, and express it in hours.
d. Chip designers would very much like to make the oxide thinner, because a thinner oxide means a smaller, cheaper, faster cell. Using your answers, explain quantitatively why they cannot, and identify the feature of the tunneling formula responsible for the wall they run into.

40.22 ★★★ *[Real-world application: chemistry — why long molecules are colored.]* The π electrons of a conjugated molecule (a chain of carbon atoms with alternating single and double bonds) are delocalized along the chain and can be modeled crudely as particles in a one-dimensional infinite well whose width L is the length of the chain. Because electrons are fermions, each level holds exactly two of them (Chapter 41). Consider a polyene with $18\ \pi$ electrons in a chain of length $L = 1.8\ \text{nm}$.

a. Find $E_1 = h^2/8mL^2$ in joules and in electron-volts.
b. Which is the highest occupied level n, and which the lowest unoccupied one?
c. Find the energy of the lowest-energy absorption (the jump from the highest occupied to the lowest unoccupied level) and the wavelength of light absorbed. What color is absorbed, and hence what color does the molecule appear?
d. A longer polyene has $22\ \pi$ electrons in a chain of $L = 2.2\ \text{nm}$. Find its absorption wavelength. Then show from $E_n \propto n^2/L^2$ that lengthening a conjugated chain always shifts the absorption toward the red, and use this to explain why short molecules like ethylene are colorless while long ones like β-carotene are vividly colored.

40.23 ★★★ *[Real-world application: nanotechnology — a quantum wire.]* An electron is confined in a rectangular box of dimensions $L_x = 5.0\ \text{nm}$, $L_y = L_z = 1.2\ \text{nm}$: long in one direction and tightly squeezed in the other two. Its energies are

$$E_{n_x,n_y,n_z} = \frac{h^2}{8m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right). \tag{40.E.2}$$

a. Find the ground-state energy $E_{1,1,1}$ in eV.
b. Find $E_{2,1,1}$ and $E_{1,2,1}$. Which costs more energy: exciting the long direction or the short one? By what factor?
c. Counting states in order of increasing energy, how many states with $n_y = n_z = 1$ lie below $E_{1,2,1}$? List their energies.
d. Your answer to (c) says that over a whole band of energies the electron has many available states along x but only one choice in y and z. Explain why the structure is therefore called a one-dimensional "quantum wire," and predict qualitatively what changes if L_y and L_z are increased to $5.0\ \text{nm}$ as well.

This page titled 40.E: A Deeper Look at Quantum Mechanics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41: ATOMIC PHYSICS

In 1868 the astronomer Norman Lockyer pointed a spectroscope at the edge of the Sun during an eclipse and found a bright yellow line that matched no element known on Earth. He proposed that it belonged to a new element and named it **helium**, after *Helios*, the Greek sun-god. For nearly thirty years helium was an element known only in the sky; not until 1895 was it finally isolated in a laboratory. How could anyone discover an element in an object $1.5 \times 10^{8}\ \text{km}$ away that they could never touch, weigh, or bottle?

The answer is that every element broadcasts a fingerprint. When the atoms of a gas are heated, they emit light at a few sharp, precise wavelengths, and no two elements share the same set. Sodium glows yellow, neon red, mercury blue-white, and hydrogen emits a familiar quartet of visible lines at 656, 486, 434, and $410\ \text{nm}$. These patterns are so reliable that astronomers read them in starlight to determine what distant stars, nebulae, and even exoplanet atmospheres are made of. The barcode of light an atom emits is dictated entirely by the discrete energies its electrons are allowed to have, and in this chapter we will compute those energies from first principles for the simplest atom, hydrogen, and see exactly why its fingerprint looks the way it does. By the end you will be able to take a set of measured wavelengths and name the gas that produced them, just as Lockyer read helium in the Sun.

Learning Objectives

- Trace how the model of the atom evolved from a solid sphere to a quantum probability cloud.
- Explain why an electron bound in an atom can have only discrete, quantized energies.
- Derive the Bohr energy levels of hydrogen, $E_n = -13.6\ \text{eV}/n^2$, and the Bohr radius.
- Use the Rydberg formula to find the wavelengths of the hydrogen spectral series and identify emission and absorption spectra.
- Write the ground-state wave function of hydrogen and compute radial probabilities such as the most probable radius and $\langle r \rangle$.
- Assign the quantum numbers n, ℓ, and m_ℓ to atomic states and interpret them physically as energy, angular-momentum magnitude, and angular-momentum direction.
- Apply electron spin and the Pauli exclusion principle to build up the electron configurations of multi-electron atoms.
- Explain absorption, spontaneous emission, and stimulated emission, and describe how a population inversion and an optical cavity make a laser.

We will return to the spectral fingerprint once we can compute hydrogen's lines, and in the final problems you will identify an unknown gas from its lines.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

41.1: Models of the Atom
41.2: Energies of a Trapped Electron in an Atom
41.3: The Hydrogen Atom
41.4: Atomic Spectra of Gases
41.5: The Quantum Model of the Hydrogen Atom
41.6: The Pauli Exclusion Principle
41.7: Lasers
41.S: Atomic Physics (Summary)

41.E: Atomic Physics (Problems)

This page titled 41: Atomic Physics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.1: Models of the Atom

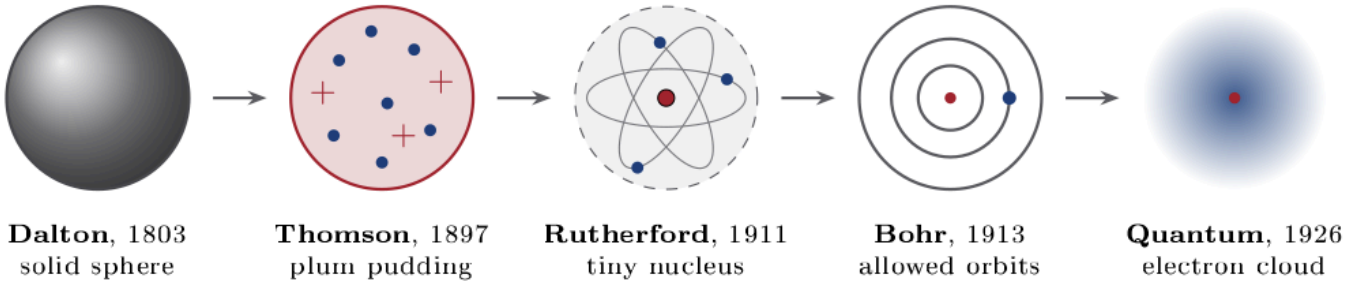


Figure 41.1. Five models of the atom. Each was overturned by an experiment it could not explain: Thomson's discovery of the electron, Rutherford's gold-foil scattering, the stability and sharp spectra of atoms, and finally the wave nature of the electron.

•Figure description

Five models of the atom in a row, joined by arrows from left to right. Dalton, 1803: a solid shaded gray sphere. Thomson, 1897: a pale red sphere of positive charge, marked with plus signs, with small blue electrons scattered through it, labeled plum pudding. Rutherford, 1911: a tiny red nucleus at the center of a mostly empty atom, outlined by a dashed circle, with electrons circling it like planets on three tilted elliptical orbits, labeled tiny nucleus. Bohr, 1913: a red nucleus surrounded by three concentric circular orbits, with one electron on the middle orbit, labeled allowed orbits. Quantum, 1926: a red nucleus inside a blue haze that is densest at the center and fades outward, labeled electron cloud.

The idea that matter is built from indivisible pieces is ancient, but the first modern *atomic theory* came from John Dalton in the early 1800s. Dalton proposed that each element is made of identical atoms that cannot be created, destroyed, or subdivided, and that compounds form when atoms of different elements combine in simple whole-number ratios. In this earliest picture (sometimes called the **billiard-ball model**) the atom was a tiny, uniform, solid sphere, the first of the five models in Figure 41.1.

That picture survived until 1897, when J. J. Thomson discovered the electron and proved that atoms are in fact *divisible*. Because bulk matter is electrically neutral, Thomson reasoned that the negative electrons must be embedded in a spread-out cloud of positive charge, like raisins in a cake. This is the **plum-pudding model**.

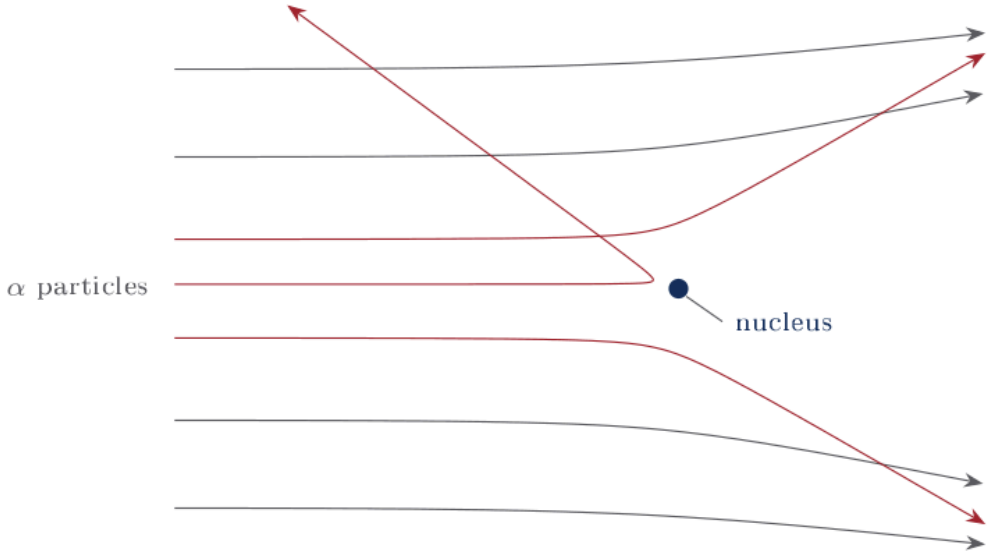


Figure 41.2. Rutherford's gold-foil experiment (schematic). Most alpha particles pass nearly straight through the mostly empty atom (gray), but those aimed close to the tiny, dense nucleus are deflected through large angles, and a near head-on particle is scattered almost straight back (red). Only a minuscule, massive, positively charged core can turn a fast alpha particle around like this, which is fatal to the plum-pudding model.

•Figure description

Rutherford's gold-foil experiment, drawn schematically. A beam of alpha particles travels from left to right toward a single small dark dot labeled nucleus. Most of the paths, drawn in gray, pass well above or below the nucleus and are bent only slightly. Two red paths that pass close to the nucleus are bent noticeably, one upward and one downward. A third red path aimed almost head-on at the nucleus is turned right around and sent back toward the upper left. Only a tiny, dense, massive core could reverse a fast alpha particle like that.

The plum pudding was overthrown by one of the most famous experiments in physics. In 1911 Ernest Rutherford's team fired energetic, positively charged alpha particles at a thin gold foil. If the positive charge were smeared out as Thomson imagined, every alpha particle should have sailed through with only a tiny deflection. Instead, while most did pass nearly straight through, a small fraction bounced back at large angles (Figure 41.2), as if they had struck something tiny, hard, and massive. Rutherford concluded that an atom's positive charge and nearly all its mass are concentrated in a minuscule central **nucleus**, with the electrons occupying the comparatively vast space around it. The atom, it turned out, is mostly empty.

Checkpoint 41.1

Most alpha particles passed straight through the gold foil, but a few rebounded at large angles. Explain in one or two sentences why this result is fatal to the plum-pudding model but is exactly what a tiny, dense nucleus predicts.

Answer

Rutherford's nuclear atom had a serious flaw, though: a classical electron orbiting a nucleus should continuously radiate energy, spiral inward, and collapse in a fraction of a nanosecond. In 1913 Niels Bohr fixed this by a bold postulate: that the electron may occupy only certain *allowed* orbits, in which it does not radiate, and that light is emitted or absorbed only when it jumps between them. The Bohr model, which we develop in Section 41.3, gets the hydrogen spectrum exactly right. Yet even it is not the final word: the modern **quantum** (or electron-cloud) model, built by Schrödinger and others, replaces the electron's sharp orbit with a smeared probability cloud, as we saw for trapped particles in Chapter 40. We will arrive at that picture by the end of the chapter.

Answers to Checkpoints

Checkpoint 41.1. In the plum-pudding model the positive charge is spread thinly through the whole atom, so its electric field is everywhere too weak to deflect a fast, heavy alpha particle by more than a tiny angle. A rebound requires an enormous force, which only a very small, dense concentration of positive charge and mass can supply. The rare large-angle scatters are the few alphas that happen to pass very close to such a nucleus; the rest miss it and pass straight through.

This page titled 41.1: Models of the Atom is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.2: Energies of a Trapped Electron in an Atom

An electron bound to an atom is a *trapped* particle, held near the nucleus by electrical attraction in much the same way the particles of Chapter 40 were held inside a well. And as we found there, confinement forces quantization: a trapped electron can exist only in certain discrete states, with certain discrete energies.

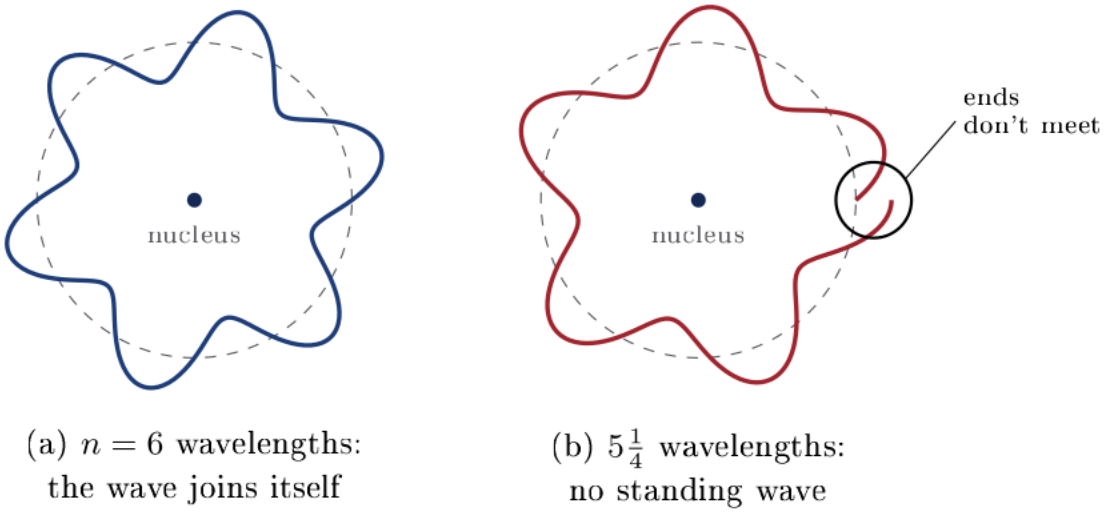


Figure 41.3. A de Broglie wave wrapped around a circular electron orbit. (a) When a whole number of wavelengths fits the circumference, $2\pi r = n\lambda$, the wave closes smoothly on itself. (b) With $5\frac{1}{4}$ wavelengths the ends do not meet, and on later trips around the wave cancels itself out. This is the same standing-wave logic that quantized the trapped particle of Chapter 40, now bent into a loop, and it is the origin of the quantum number n.

•Figure description

Two de Broglie waves wrapped around a circular electron orbit, side by side. In each, a small dot at the center is the nucleus and a dashed gray circle marks the orbit radius. Panel a: a blue wavy closed loop whose distance from the center ripples in and out six times as it goes around, six full wavelengths, labeled n equals 6; it joins up smoothly with itself, the condition 2 pi r equals n lambda. Panel b: a red wavy loop with five and a quarter wavelengths; when it comes back around, its end does not meet its start, and a black circle labeled ends don't meet marks the mismatch. Such a wave cannot form a standing wave.

The reason, once again, is the electron's wave nature. Its de Broglie wave must fit the available space as a **standing wave**, just as the wave on a guitar string of fixed length can vibrate only at its fundamental and overtones. For an electron circling a nucleus, the natural version of this condition is that a whole number of wavelengths must fit around the circumference of the orbit (Figure 41.3):

$$2\pi r = n\lambda, \qquad n = 1, 2, 3, \ldots \tag{41.1}$$

If the wave did not close on itself smoothly after one trip around, it would interfere destructively with itself on the next trip and cancel out. Only the closed, whole-number patterns survive, and, through $\lambda = h/p$, this restricts the electron's momentum and hence its energy to a discrete ladder of values.

This circular standing-wave condition was Bohr's route to the allowed orbits. Combined with the physics of the electrical force, it yields the allowed energies, which we now compute.

This page titled 41.2: Energies of a Trapped Electron in an Atom is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.3: The Hydrogen Atom

Hydrogen, with a single electron orbiting a single proton, is the simplest atom and the one we can solve exactly. Treating it as an electron trapped by the Coulomb attraction leads to the Bohr model. The electron's potential energy in the field of the proton is

$$U(r) = -\frac{e^2}{4\pi\epsilon_0 r} = -\frac{k_e e^2}{r}, \tag{41.2}$$

where $k_e = 1/4\pi\epsilon_0 = 8.99 \times 10^9\ \mathrm{N \cdot m^2/C^2}$. The minus sign signals an *attractive* force: the electron is bound, and we must do work to pull it away. Its total energy is kinetic plus potential,

$$E = K + U = \tfrac{1}{2}mv^2 - \frac{e^2}{4\pi\epsilon_0 r}. \tag{41.3}$$

Three physical ideas pin down the orbit. First, the Coulomb attraction supplies the centripetal force that holds the electron in its circular path:

$$\frac{mv^2}{r} = \frac{e^2}{4\pi\epsilon_0 r^2}. \tag{41.4}$$

Second, Bohr's quantization of the standing wave (Equation (41.1)) is equivalent to requiring the orbital angular momentum to come in whole-number multiples of $\hbar$:

$$L = mvr = n\hbar \quad \Longrightarrow \quad v = \frac{n\hbar}{mr}. \tag{41.5}$$

Multiplying Equation (41.4) by $r/2$ shows that the kinetic energy is $K = \frac{1}{2}mv^2 = e^2/(8\pi\epsilon_0 r)$, exactly half the size of the potential energy, so the total energy is

$$E = K + U = -\frac{e^2}{8\pi\epsilon_0 r}. \tag{41.6}$$

Solving Equations (41.4) and (41.5) together for r gives the allowed radii, $r_n = n^2 a_0$ (you will carry out this step in the problems), and substituting them into Equation (41.6) gives the celebrated result for the allowed energies:

Key Equation: Energy levels of hydrogen

$$E_n = -\frac{me^4}{8\epsilon_0^2 h^2}\,\frac{1}{n^2} = \frac{-2.18 \times 10^{-18}\,\mathrm{J}}{n^2} = \frac{-13.6\,\mathrm{eV}}{n^2}, \qquad n = 1, 2, 3, \ldots \tag{41.7}$$

The energies are negative because the electron is bound; the deepest level, $n = 1$, is the **ground state** at $-13.6\ \mathrm{eV}$, and the levels crowd toward $E = 0$ (a freed electron) as $n \to \infty$. The smallest orbit, $n = 1$, has radius equal to the **Bohr radius**

$$a_0 = \frac{\epsilon_0 h^2}{\pi m e^2} = 5.29 \times 10^{-11}\,\mathrm{m} = 52.9\,\mathrm{pm}, \tag{41.8}$$

a convenient natural unit for atomic sizes.

Checkpoint 41.2

The Bohr model pictures the electron as a tiny planet on a sharp circular orbit. It gives the energy levels of Equation (41.7) exactly, yet the modern quantum model abandons the idea of a definite orbit altogether. Suggest one reason the planetary picture cannot be literally correct, even though its energies are right.

Answer

When the electron drops from a higher level to a lower one, the lost energy leaves as a single photon, with

$$\Delta E = E_{\text{high}} - E_{\text{low}} = hf = \frac{hc}{\lambda}. \tag{41.9}$$

Combining this with the energy ladder (41.7) produces a compact formula for the emitted (or absorbed) wavelengths, in terms of the **Rydberg constant** $R = 1.097 \times 10^{7}\ \text{m}^{-1}$:

Key Equation: The Rydberg formula

$$\frac{1}{\lambda} = R\left(\frac{1}{n_{\text{low}}^2} - \frac{1}{n_{\text{high}}^2}\right), \qquad R = \frac{me^4}{8\epsilon_0^2 h^3 c} = 1.097373 \times 10^7\,\text{m}^{-1}. \tag{41.10}$$

Example 41.1: Photons from hydrogen, near and far

(a) The electron in a hydrogen atom drops from the $n = 2$ level to the ground level. Find the wavelength and frequency of the emitted photon.

(b) In interstellar space, highly excited "Rydberg" hydrogen atoms have been detected. Find the wavelength radio astronomers must tune to in order to catch the photon from an electron falling from $n = 273$ to $n = 272$.

Solution

(a) With $n_{\text{low}} = 1$ and $n_{\text{high}} = 2$, the Rydberg formula gives

$$\frac{1}{\lambda} = R\left(\frac{1}{1^2} - \frac{1}{2^2}\right) = (1.097 \times 10^7)(0.750) = 8.23 \times 10^6\,\text{m}^{-1}, \tag{41.3.1}$$

so $\lambda = 1.215 \times 10^{-7}\ \text{m} = \boxed{121.5\,\text{nm}}$, deep in the ultraviolet (the Lyman-α line). Its frequency is

$$f = \frac{c}{\lambda} = \frac{3.00 \times 10^8}{1.215 \times 10^{-7}} = \boxed{2.47 \times 10^{15}\,\text{Hz}}. \tag{41.3.2}$$

(b) Now $n_{\text{low}} = 272$ and $n_{\text{high}} = 273$. The two terms nearly cancel:

$$\frac{1}{\lambda} = R\left(\frac{1}{272^2} - \frac{1}{273^2}\right) = (1.097 \times 10^7)(9.88 \times 10^{-8}) = 1.08\,\text{m}^{-1}, \tag{41.3.3}$$

giving $\lambda = \boxed{0.92\,\text{m}}$: a radio wave about a meter long. Transitions between very high, very closely spaced levels emit very low-energy photons, which is why such atoms are studied with radio telescopes rather than optical ones.

Energy Level Transitions

Equation (41.9) works in both directions, and the two directions produce the two kinds of spectra.

In an **emission** process the atom starts in a higher level and falls to a lower one, releasing a photon whose energy equals the gap between the levels. This is an *inward*, downward transition, and it produces bright lines: light appearing at exactly the transition wavelengths.

In an **absorption** process the reverse happens: an atom in a lower level is struck by a photon whose energy exactly matches a gap, absorbs it, and jumps *outward* to a higher level. A photon that does not match any gap simply passes by. So when white light shines through a cool gas, the gas removes precisely the wavelengths it would otherwise emit, leaving dark lines in an otherwise continuous spectrum.

The crucial point is that emission and absorption for a given atom occur at *the same set of wavelengths*, because they involve the same energy-level gaps. A hydrogen atom emits a bright line at $656\ \text{nm}$ and absorbs a dark line at $656\ \text{nm}$; both correspond to the $n = 3 \leftrightarrow n = 2$ gap.

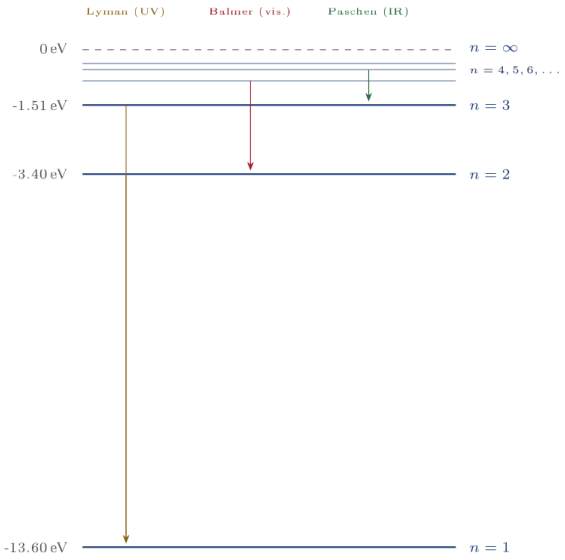


Figure 41.4. Energy-level diagram of hydrogen, with levels drawn at their energies $E_n = -13.6\ \text{eV}/n^2$ (note how they crowd toward $E = 0$). Each downward transition emits a photon; transitions ending on $n = 1$, 2, and 3 form the Lyman, Balmer, and Paschen series, lying in the ultraviolet, visible, and infrared respectively.

•Figure description

An energy-level diagram for hydrogen. Horizontal blue lines mark the allowed energies, drawn at their true spacing: n equals 1 at the bottom at minus 13.60 electron volts, n equals 2 far above it at minus 3.40, n equals 3 at minus 1.51, then several fainter unlabeled lines for n equals 4, 5, 6 and beyond crowding closer and closer together toward a dashed line at 0 electron volts marked n equals infinity. Three groups of downward arrows show transitions: gold arrows landing on n equals 1 are the Lyman series in the ultraviolet, red arrows landing on n equals 2 are the Balmer series in the visible, and green arrows landing on n equals 3 are the Paschen series in the infrared.

Example 41.2: The red line of hydrogen

Find the wavelength of the photon emitted when a hydrogen electron drops from $n = 3$ to $n = 2$, and identify its color.

Solution

The energy gap is

$$\Delta E = E_3 - E_2 = -13.6\left(\frac{1}{9} - \frac{1}{4}\right)\text{eV} = 13.6\left(\frac{1}{4} - \frac{1}{9}\right)\text{eV} = 1.89\,\text{eV}. \tag{41.3.4}$$

Using $hc = 1240\ \text{eV}\cdot\text{nm}$,

$$\lambda = \frac{hc}{\Delta E} = \frac{1240}{1.89} = \boxed{656\,\text{nm}}, \tag{41.3.5}$$

a red photon. This is the H-α line, the brightest member of hydrogen's visible series and the source of the rosy glow of many nebulae.

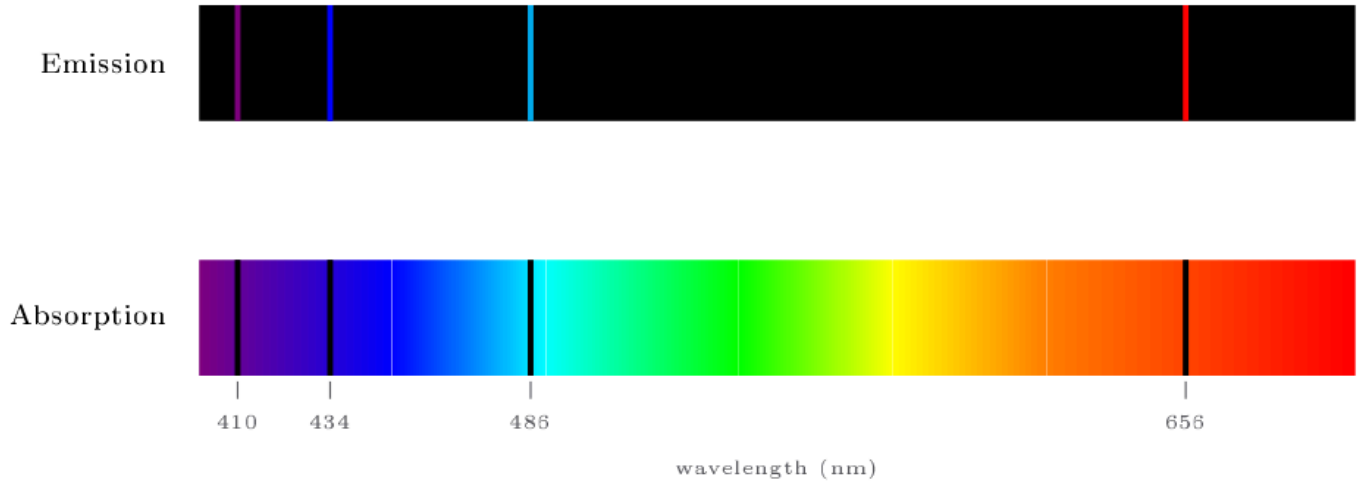


Figure 41.5. The visible (Balmer) spectrum of hydrogen: its spectral fingerprint. *Top:* an emission spectrum shows four bright lines on a dark background at 410, 434, 486, and 656 nm. *Bottom:* an absorption spectrum shows dark lines at exactly the same wavelengths, cut out of an otherwise continuous spectrum. The pattern is unique to hydrogen, and matching it is how the gas is identified.

•Figure description

The visible spectrum of hydrogen shown two ways. The upper bar is an emission spectrum: a black background carrying four bright colored lines: violet at 410 nanometers, blue at 434, cyan at 486, and red at 656, the red one sitting well apart from the other three, which cluster at the blue end. The lower bar is an absorption spectrum: a continuous rainbow running violet through red, interrupted by four black lines at exactly those same four wavelengths. The same energy gaps produce bright lines when the gas emits and dark lines when it absorbs.

Answers to Checkpoints

Checkpoint 41.2. An electron circling on a sharp orbit is constantly accelerating, and classical electromagnetism says an accelerating charge radiates. It would lose energy and spiral into the nucleus in a tiny fraction of a second, so the atom could not be stable. A second reason: a sharp orbit would give the electron a definite position and momentum at once, which the uncertainty principle forbids. (A third: in the ground state the real atom has zero orbital angular momentum, not the $\hbar$ Bohr assigns it.)

This page titled 41.3: The Hydrogen Atom is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.4: Atomic Spectra of Gases

Because the lower level n_{low} of a transition can be the ground state or any excited state, hydrogen's lines fall into families called **series**, each defined by where the electron lands (Figure 41.4). Transitions ending on $n = 1$ form the **Lyman series**, all in the ultraviolet. Transitions ending on $n = 2$ form the **Balmer series**, four of whose lines (at 656.3, 486.1, 434.0, and $410.1\ \text{nm}$) lie in the visible and give hydrogen its characteristic look (Figure 41.5). Transitions ending on $n = 3$ form the **Paschen series**, in the infrared.

The four visible Balmer lines are what an astronomer actually records, and together they form the recognizable pattern shown in Figure 41.5: bright lines at those wavelengths if the gas is glowing (emission), or dark lines at exactly the same wavelengths if a continuous source shines through cooler gas (absorption).

Every element has its own ladder of levels and therefore its own unique pattern of lines: a **spectral fingerprint**. This is what makes spectroscopy such a powerful tool: by photographing the light from a glowing gas and matching its line pattern against known elements, a scientist can identify the gas without ever sampling it. Astronomers use exactly this method to determine the composition of stars, nebulae, and exoplanet atmospheres light-years away, and it is how helium was found in the Sun decades before it was found on Earth.

Checkpoint 41.3

(a) A hydrogen atom's electron drops from $n = 5$ to $n = 2$. Using Equation (41.7), find the photon's energy and wavelength. What color is it? (b) A hydrogen atom in its ground state is struck by a $10.0\ \text{eV}$ photon. Can it absorb the photon?

Answer

Answers to Checkpoints

Checkpoint 41.3. (a) $E = 13.6\ \text{eV}\left(\frac{1}{2^2} - \frac{1}{5^2}\right) = 13.6\left(\frac{1}{4} - \frac{1}{25}\right) = 2.86\ \text{eV}$, and $\lambda = hc/E = (1240\ \text{eV}\cdot\text{nm})/(2.86\ \text{eV}) \approx 434\ \text{nm}$: violet light, the third of hydrogen's four visible lines. (b) No. The smallest jump from the ground state, $n = 1 \to 2$, needs exactly $13.6\left(1 - \frac{1}{4}\right) = 10.2\ \text{eV}$. An atom absorbs a photon only if its energy matches the gap between two levels, and $10.0\ \text{eV}$ matches none, so the photon passes by.

This page titled 41.4: Atomic Spectra of Gases is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.5: The Quantum Model of the Hydrogen Atom

The Bohr model gets the energies right but clings to the classical fiction of a definite orbit. The full treatment, due to Schrödinger, drops that fiction and solves the three-dimensional time-independent Schrödinger equation for the electron in the proton's Coulomb potential:

$$-\frac{\hbar^2}{2m}\left(\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+\frac{\mathrm{d}^2\psi}{\mathrm{d}y^2}+\frac{\mathrm{d}^2\psi}{\mathrm{d}z^2}\right)-\frac{k_e e^2}{r}\,\psi = E\psi. \tag{41.11}$$

Solving and normalizing gives, for the ground state, a wave function that depends only on the distance r from the nucleus:

Key Equation: Hydrogen ground-state wave function

$$\psi_{1s}(r) = \frac{1}{\sqrt{\pi}\,a_0^{3/2}}\,e^{-r/a_0}, \tag{41.12}$$

where a_0 is again the Bohr radius. Remarkably, the Schrödinger treatment reproduces the very same energy levels (41.7) that Bohr obtained, but now the electron is not on an orbit at all. It is spread out as a probability cloud, densest near the nucleus and thinning out with distance.

To turn the wave function into the probability of finding the electron somewhere, we integrate $|\psi|^2$ over a volume. Because ψ_{1s} is spherically symmetric, the natural volume element is a thin spherical shell of radius r and thickness dr, whose volume is $dV = 4\pi r^2\,dr$. The probability of finding the electron in that shell defines the **radial probability density** $P(r)$:

Key Equation: Radial probability density of the $1s$ state

$$P(r)\,dr = |\psi_{1s}|^2\,4\pi r^2\,dr \quad\Longrightarrow\quad P(r) = \frac{4r^2}{a_0^3}\,e^{-2r/a_0}. \tag{41.13}$$

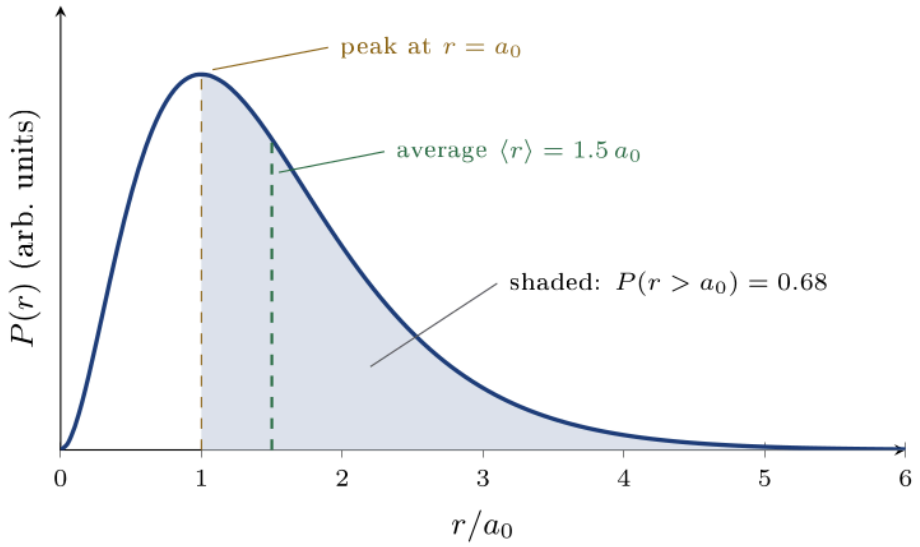


Figure 41.6. Radial probability density $P(r) = (4r^2/a_0^3)\ e^{-2r/a_0}$ for the hydrogen ground state. It vanishes at the nucleus, rises to a maximum at exactly the Bohr radius $r = a_0$, and falls off exponentially beyond. The electron is most likely to be found near a_0, but the long tail gives a 68% chance of finding it beyond a_0 (shaded) and pulls the average radius out to $\langle r \rangle = 1.5\ a_0$ (Example 41.3).

•Figure description

A graph of the radial probability density P of r for the hydrogen ground state, plotted against radius in units of the Bohr radius a nought. The curve starts at exactly zero at the nucleus, rises steeply to a single maximum at r equals one Bohr radius, marked by a dashed line, then falls away exponentially, still noticeably above zero at two and three Bohr radii and fading toward nothing by about six. The electron is most likely found at a nought, but has a long tail of probability much farther out. The area under the curve beyond a nought is shaded and labeled P of r greater than a nought equals 0.68, and a green dashed line at 1.5 a nought marks the average radius.

The factor r^2 pulls $P(r)$ to zero at the nucleus (a thin shell there has almost no volume), while the exponential pulls it down at large r; in between, $P(r)$ peaks (Figure 41.6). That peak, as the next example shows, sits exactly at the Bohr radius.

Example 41.3: Where is the ground-state electron?

For an electron in the hydrogen ground state, **(a)** find the most probable distance from the nucleus; **(b)** find the probability of finding it *outside* the Bohr radius; and **(c)** find its average distance $\langle r \rangle$.

Solution

(a) The most probable radius maximizes $P(r) \propto r^2 e^{-2r/a_0}$. Setting the derivative to zero,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2 e^{-2r/a_0}\right) = \left(2r - \frac{2r^2}{a_0}\right) e^{-2r/a_0} = 0 \implies r = \boxed{a_0 = 52.9\,\mathrm{pm}}. \tag{41.5.1}$$

The probability cloud is densest at exactly the radius of Bohr's first orbit: the two models agree on this length even though they disagree on what the electron is doing there.

(b) Letting $u = r/a_0$, the probability of finding the electron beyond a_0 is

$$P(r > a_0) = \int_{a_0}^{\infty} \frac{4r^2}{a_0^3} e^{-2r/a_0}\, dr = \int_1^{\infty} 4u^2 e^{-2u}\, du = 5e^{-2} = \boxed{0.677}. \tag{41.5.2}$$

More than two-thirds of the time the electron is farther from the nucleus than Bohr's orbit would allow: a purely quantum spreading with no classical counterpart.

(c) The average radius is

$$\langle r\rangle = \int_0^\infty r\,P(r)\,dr = \frac{4}{a_0^3}\int_0^\infty r^3 e^{-2r/a_0}\,dr = \frac{4}{a_0^3}\cdot\frac{6}{(2/a_0)^4} = \boxed{1.5\,a_0 = 79.4\,\text{pm}}. \tag{41.5.3}$$

The average exceeds the most probable value because the long exponential tail at large r skews the distribution outward.

The Wave Functions of a Hydrogen Atom

The ground state is spherically symmetric, but excited states are not. In general the electron's position depends not only on the radius r but on the two angles θ and ϕ that locate it on a sphere. The three-dimensional Schrödinger equation can be separated into three ordinary differential equations (one for each coordinate) so that the full wave function factors as

$$\psi(r,\theta,\phi) = R(r)\,f(\theta)\,g(\phi). \tag{41.14}$$

Each factor must satisfy its own boundary conditions, and each therefore introduces its own **quantum number**. The radial factor $R(r)$ gives the *principal* quantum number n; the polar factor $f(\theta)$ gives the *orbital* quantum number ℓ; and the azimuthal factor $g(\phi)$ gives the *orbital magnetic* quantum number m_ℓ. Their allowed values are linked:

Key Equation: Allowed quantum numbers

$$n = 1, 2, 3, \ldots; \qquad \ell = 0, 1, 2, \ldots, n-1; \qquad m_\ell = -\ell, -\ell+1, \ldots, +\ell. \tag{41.15}$$

So for a given n there are n possible values of ℓ, and for each ℓ there are $2\ell+1$ possible values of m_ℓ. (For the ground state these extra quantum numbers have no choice: $n=1$ forces $\ell=0$ and $m_\ell=0$.)

For historical reasons, all states sharing the same n form a **shell**, labeled K, L, M, ... for $n = 1, 2, 3, \ldots$. All states sharing the same n and ℓ form a **subshell**, labeled by a letter:

$$\ell = 0, 1, 2, 3, 4, \ldots \;\longleftrightarrow\; s, p, d, f, g, \ldots \tag{41.5.4}$$

Thus the state with $n=3, \ell=1$ is called $3p$, and $n=2, \ell=0$ is $2s$. The shapes of the corresponding probability clouds become more intricate as ℓ grows: s states ($\ell=0$) are spherical, p states ($\ell=1$) are dumbbell-shaped, most d states ($\ell=2$) have four lobes, and so on (Figure 41.7).

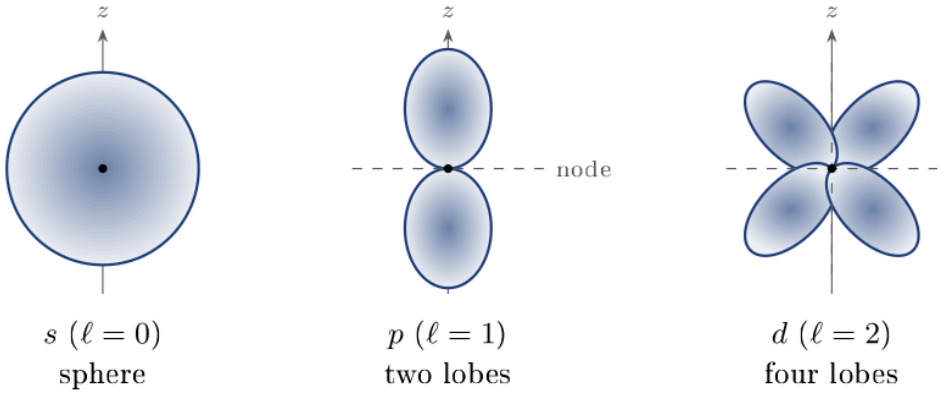


Figure 41.7. The shapes of s, p, and d probability clouds. Dashed lines are nodes, where the probability is zero; a state with orbital quantum number ℓ has ℓ such angular nodes. (These are sketches of where the electron is likely to be found, not sharp surfaces. The p cloud shown has $m_\ell = 0$; the four-lobed d cloud is a mixture of $m_\ell = +1$ and -1, as explained with Figure 41.9.)

•Figure description

Sketches of three hydrogen probability clouds, each with a vertical z axis. The s cloud, for l equals 0, is a sphere, densest at the center. The p cloud, for l equals 1, is two egg-shaped lobes along the z axis, one above and one below the nucleus, with a dashed horizontal line through the nucleus marking a node where the probability is zero. The d cloud, for l equals 2, is four lobes arranged like a four-leaf clover between the axes, with dashed horizontal and vertical lines through the nucleus marking two nodes.

Checkpoint 41.4

For a given shell n, why is the largest allowed value of ℓ equal to $n-1$ rather than n? (Think about how many total nodes a state of principal quantum number n is allowed, split between radial and angular parts.)

Answer

Example 41.4: Counting the states in a shell

How many distinct (n, ℓ, m_ℓ) states does the $n = 3$ shell contain?

Solution

With $n = 3$, the allowed orbital quantum numbers are $\ell = 0, 1, 2$, and each contributes $2\ell + 1$ values of m_ℓ:

$$\ell = 0:\ 1 \text{ state}, \quad \ell = 1:\ 3 \text{ states}, \quad \ell = 2:\ 5 \text{ states}. \tag{41.5.5}$$

The total is $1 + 3 + 5 = \boxed{9}$ orbital states, which equals n^2. (As we will see in Section 41.6, electron spin doubles this to $2n^2 = 18$ available electron states.)

Physical Interpretation of Quantum Numbers

The three quantum numbers are not just bookkeeping labels; each controls a measurable physical quantity. The key is that *energy is a scalar* but *angular momentum is a vector*, requiring both a magnitude and a direction to specify.

The **principal** quantum number n fixes the energy, exactly as in the Bohr model:

$$E_n = \frac{-13.6\,\text{eV}}{n^2}. \tag{41.16}$$

The **orbital** quantum number ℓ fixes the *magnitude* of the electron's orbital angular momentum, while the **orbital magnetic** quantum number m_ℓ fixes its *direction* through the component L_z along a chosen axis:

Key Equation: Magnitude and direction of orbital angular momentum

$$L = \sqrt{\ell(\ell+1)}\,\hbar, \qquad L_z = m_\ell\,\hbar. \tag{41.17}$$

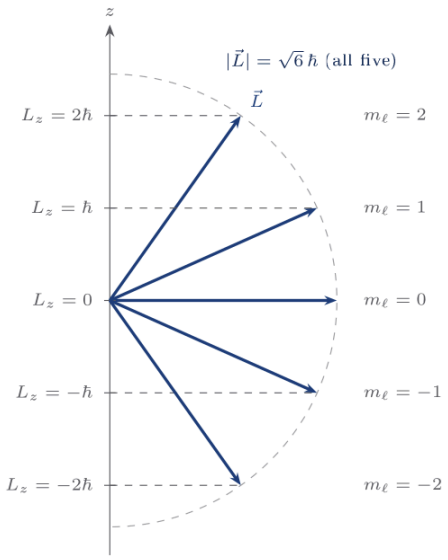


Figure 41.8. Space quantization for $\ell = 2$. The angular-momentum vector has fixed magnitude $|\mathbf{L}| = \sqrt{\ell(\ell+1)}\ \hbar = \sqrt{6}\ \hbar$, but its projection on the z-axis is restricted to $L_z = m_\ell \hbar$ with $m_\ell = -2, \dots, +2$. The vector therefore lies on one of five discrete cones; it can never align fully with the axis.

•Figure description

Space quantization for an orbital quantum number of 2, drawn as five arrows radiating from a common origin against a vertical z axis. All five arrows have the same length, the fixed magnitude of L equal to root six times h-bar, but each points at a different angle, so their projections onto the z axis take the five values m sub l equals plus 2, plus 1, 0, minus 1 and minus 2 times h-bar, marked by dashed horizontal lines. Even the steepest arrow, m sub l equals plus 2, leans well away from the axis: the vector can never align fully with z.

Because m_ℓ takes only the integer values from $-\ell$ to $+\ell$, the vector $\mathbf{L}$ can point only along certain discrete cones around the z-axis (usually set by an external magnetic field), a phenomenon called **space quantization** (Figure 41.8). The angle θ between $\mathbf{L}$ and the z-axis is set by

$$\cos\theta = \frac{L_z}{L} = \frac{m_\ell}{\sqrt{\ell(\ell+1)}}. \tag{41.18}$$

Notice that L_z is always strictly less than L (since $|m_\ell| \le \ell < \sqrt{\ell(\ell+1)}$): the angular-momentum vector can never point exactly along the axis, because that would mean its x- and y-components were both exactly zero, violating the uncertainty principle. Instead $\mathbf{L}$ lies somewhere on the surface of a cone.

The orbital magnetic quantum number shapes the probability cloud as well as the direction of $\mathbf{L}$. For a state with a definite m_ℓ, the probability density does not depend on the azimuthal angle ϕ, so every such cloud is symmetric about the z-axis: a cross-section spun around z (Figure 41.9). The larger $|m_\ell|$ is, the closer $\mathbf{L}$ lies to the z-axis, and the more the electron circulates *around* that axis rather than along it. So an $m_\ell = 0$ cloud stretches along z, while an $|m_\ell| = \ell$ cloud is pushed out into a doughnut around the equator. States with $+m_\ell$ and $-m_\ell$ have identical clouds; they differ only in the direction the electron circulates. (The p lobes along x and y familiar from chemistry, and the four-lobed d cloud of Figure 41.7, are equal mixtures of a $+m_\ell$ state and a $-m_\ell$ state. They are perfectly good states, but not ones with a definite L_z.)

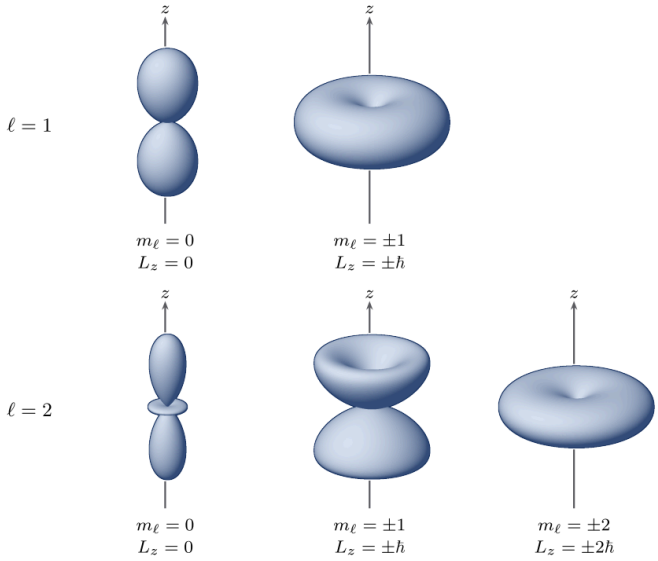


Figure 41.9. Hydrogen probability clouds for each value of m_ℓ. Every cloud is symmetric about the z-axis. As $|m_\ell|$ grows, $\mathbf{L}$ tips toward the z-axis (Figure 41.8) and the cloud is pushed away from the axis toward the equator, so $m_\ell = \pm\ell$ gives a doughnut. A state and its partner with the opposite sign of m_ℓ have the same cloud.

•Figure description

Three-dimensional probability clouds for each allowed value of m sub l, each drawn as a shaded blue solid around a vertical z axis. Top row, l equals 1: for m sub l equals 0, L sub z equals 0, two egg-shaped lobes stacked along the z axis, one above and one below the nucleus. For m sub l equals plus or minus 1, L sub z equals plus or minus h-bar, a fat doughnut lying flat around the z axis, with a dimple at the center where the axis passes through. Bottom row, l equals 2: for m sub l equals 0, two long lobes along z with a thin ring around the equator. For m sub l equals plus or minus 1, two bowls, one above and one below the equator, each opening away from the nucleus, like two cones of probability. For m sub l equals plus or minus 2, L sub z equals plus or minus 2 h-bar, a flatter doughnut around the equator. As the size of m sub l grows, the cloud moves away from the z axis toward the equator.

Example 41.5: Angular momentum of an $\ell = 3$ state

A hydrogen atom is in a state with $\ell = 3$. Find the magnitude of $\mathbf{L}$, the allowed values of L_z, and the angles $\mathbf{L}$ makes with the z-axis.

Solution

The magnitude is fixed by ℓ:

$$L = \sqrt{\ell(\ell+1)}\,\hbar = \sqrt{12}\,\hbar = \boxed{3.65 \times 10^{-34}\,\mathrm{J \cdot s}}. \tag{41.5.6}$$

The allowed projections are $L_z = m_\ell\hbar$ for the seven values $m_\ell = -3, -2, -1, 0, 1, 2, 3$, i.e. L_z ranges from $-3\hbar$ to $+3\hbar$ in steps of $\hbar$. The corresponding angles follow from $\cos\theta = m_\ell/\sqrt{12}$:

$$\begin{aligned} m_\ell = 3 &\to \theta = 30.0^\circ, & m_\ell = 2 &\to 54.7^\circ, & m_\ell = 1 &\to 73.2^\circ, \\ m_\ell = 0 &\to 90.0^\circ, & m_\ell = -1 &\to 106.8^\circ, & m_\ell = -2 &\to 125.3^\circ, \end{aligned} \tag{41.5.7}$$

and $m_\ell = -3 \to \boxed{150.0^\circ}$. The seven cones are symmetric about the equatorial (90°) plane, and none reaches the poles.

This discreteness also illuminates the boundary between the quantum and classical worlds. For a state with a very large principal quantum number, say $n = 45$ and $\ell = n - 1 = 44$, the probability cloud bunches into a thin ring of radius $r = n^2 a_0$, and the closely spaced cones blur into a continuum. The electron then behaves almost like a classical particle on a circular orbit, in keeping with Bohr's *correspondence principle*: quantum predictions must merge into classical ones in the limit of large quantum numbers.

Answers to Checkpoints

Checkpoint 41.4. A state with principal quantum number n has $n-1$ nodes in all, shared between the radial part and the angular part; the angular part has ℓ of them. The number of radial nodes, $n-1-\ell$, cannot be negative, so ℓ can be at most $n-1$. A state with $\ell = n$ would need a negative number of radial nodes.

This page titled 41.5: The Quantum Model of the Hydrogen Atom is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.6: The Pauli Exclusion Principle

So far three quantum numbers have described the electron's spatial state. But experiment (notably the splitting of spectral lines in a magnetic field) reveals a fourth, with no classical analog: the electron carries an intrinsic angular momentum, or **spin**, as if it were a tiny spinning charge. Spin is quantized just as orbital angular momentum is, but for an electron it allows only two orientations, described by the **spin magnetic quantum number**

$$m_s = +\tfrac{1}{2} \quad \text{or} \quad m_s = -\tfrac{1}{2}, \tag{41.19}$$

informally called "spin up" and "spin down." A complete specification of an electron's state in an atom therefore requires four quantum numbers: (n, ℓ, m_ℓ, m_s).

The reason this matters for the structure of atoms is a deep principle discovered by Wolfgang Pauli:

Key Equation: The Pauli exclusion principle

No two electrons in the same atom can occupy the same quantum state, that is, no two electrons can share the same set of all four quantum numbers (n, ℓ, m_ℓ, m_s).

The consequences are enormous. Without the exclusion principle, every electron in an atom would simply tumble into the lowest-energy $1s$ state, and all elements would behave almost alike. Instead, electrons are forced to stack up into successively higher states, and the resulting layered structure is the origin of chemistry and the periodic table.

Counting the available states shows how many electrons each level can hold. Each (n, ℓ, m_ℓ) orbital can hold exactly two electrons, one of each spin. A subshell with orbital quantum number ℓ has $2\ell + 1$ orbitals and so holds $2(2\ell + 1)$ electrons: an s subshell holds 2, a p holds 6, a d holds 10, and an f holds 14. A full shell n holds

$$2\sum_{\ell=0}^{n-1}(2\ell + 1) = 2n^2 \tag{41.20}$$

electrons, so the K, L, and M shells hold 2, 8, and 18 electrons respectively. Filling these states from the bottom up, lowest energy first, gives an atom's **ground-state electron configuration**.

Example 41.6: Building up an atom

Sodium has atomic number $Z = 11$. Write its ground-state electron configuration and explain what it implies about sodium's chemistry.

Solution

We place eleven electrons into the lowest available states, respecting the capacities $2(2\ell + 1)$ of each subshell (Figure 41.10):

$$\underbrace{1s^2}_{2}\ \underbrace{2s^2\,2p^6}_{8}\ \underbrace{3s^1}_{1} \qquad (2 + 2 + 6 + 1 = 11). \tag{41.6.1}$$

The first two shells are completely filled (a stable, inert core of ten electrons), leaving a single **valence electron** alone in the $3s$ subshell. Because that lone outer electron is easily removed, sodium is highly reactive and forms $+1$ ions, exactly the behavior of the alkali metals in the first column of the periodic table.

$$\boxed{1s^2\,2s^2\,2p^6\,3s^1} \tag{41.6.2}$$

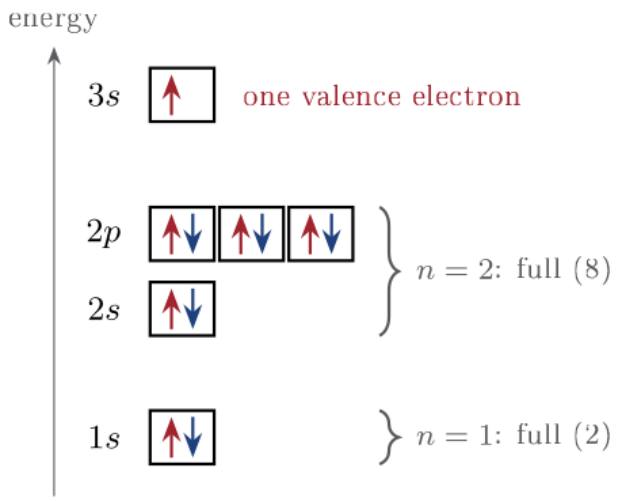


Figure 41.10. The ground state of sodium ($Z = 11$), $1s^2\ 2s^2\ 2p^6\ 3s^1$. Each box is one orbital (n, ℓ, m_ℓ); the exclusion principle allows it at most two electrons, one with $m_s = +\frac{1}{2}$ (up) and one with $m_s = -\frac{1}{2}$ (down). The filled $n = 1$ and $n = 2$ shells form an inert core; the lone $3s$ electron is the valence electron that sets sodium's chemistry.

•Figure description

An orbital-filling diagram for the ground state of sodium, with energy increasing upward. Each box is one orbital, a single set of n, l and m sub l, and holds at most two electrons, drawn as one up arrow and one down arrow for the two spin states. From the bottom: the 1s box holds two electrons; the 2s box holds two; the three 2p boxes, side by side at slightly higher energy, hold two each; and at the top the single 3s box holds one lone up arrow, sodium's valence electron. Braces group the full n equals 1 and n equals 2 shells as the inert core of ten electrons.

The periodic table, then, is not an arbitrary chart but a direct map of how quantum states fill under the exclusion principle. The lengths of its rows $(2, 8, 8, 18, \ldots)$ are the combined capacities of the subshells filled across each row: 2 for $1s$, then 8 for $2s$ and $2p$, then 8 for $3s$ and $3p$, and so on. Elements in the same column share the same outer-electron arrangement, and hence similar chemistry. The behavior of every element traces back to the four quantum numbers and Pauli's single, powerful rule.

Checkpoint 41.5

(a) How many electrons can occupy the $n = 2$ shell? Count the allowed sets of quantum numbers. (b) Without the exclusion principle, where would all six of carbon's electrons be in its ground state, and why would chemistry be very different?

Answer

Answers to Checkpoints

Checkpoint 41.5. (a) Eight. For $n = 2$: $\ell = 0$ allows $m_\ell = 0$, and $\ell = 1$ allows $m_\ell = -1, 0, +1$, giving four spatial states; each can hold two electrons, spin up and spin down. That makes 8, in agreement with the general rule $2n^2$. (b) All six electrons would crowd into the lowest-energy state, the $1s$ orbital. Every element would then have essentially the same compact, tightly bound electron arrangement, with no outer shells of different sizes to fill. The periodic table's pattern of chemical behavior (which comes from the partly filled outer shells) would vanish.

This page titled 41.6: The Pauli Exclusion Principle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.7: Lasers

Every light source we have met so far (a hot filament, a glowing gas, the Sun) emits light the same way: each atom, once excited, drops to a lower level on its own and releases a photon in a random direction at a random moment. The result is a jumble of independent wavetrains. A **laser** is different in kind. Its light is a single, orderly wave: one wavelength, marching in step, in one direction. The name is an acronym for what makes this possible: *light amplification by stimulated emission of radiation*.

Three ways light meets an atom. Consider two levels of an atom, E_1 below and E_2 above, and light whose photons carry exactly the difference, $hf = E_2 - E_1$. Three processes can occur (Figure 41.11).

- **Absorption**: an atom in the lower level absorbs a photon and jumps to the upper level. The photon disappears.
- **Spontaneous emission**: an atom in the upper level drops to the lower level on its own, emitting a photon in a random direction with a random phase. A typical excited state lasts about $10\ \text{ns}$ before this happens.
- **Stimulated emission**: a passing photon of energy hf triggers an atom *already in* the upper level to drop, and the emitted photon is an exact copy of the one that triggered it: the same frequency, the same direction, the same polarization, and in phase with it. One photon goes in; two identical photons come out.

Einstein predicted stimulated emission in 1917 from a thermodynamic argument: without it, atoms bathed in thermal radiation could not reach the equilibrium that the blackbody spectrum (Chapter 39) requires. He also showed that a photon is exactly as likely to stimulate emission from an excited atom as it is to be absorbed by an unexcited one. That symmetry decides everything. Light passing through a collection of atoms is amplified if more of them are in the upper level than in the lower, and weakened if the reverse is true.

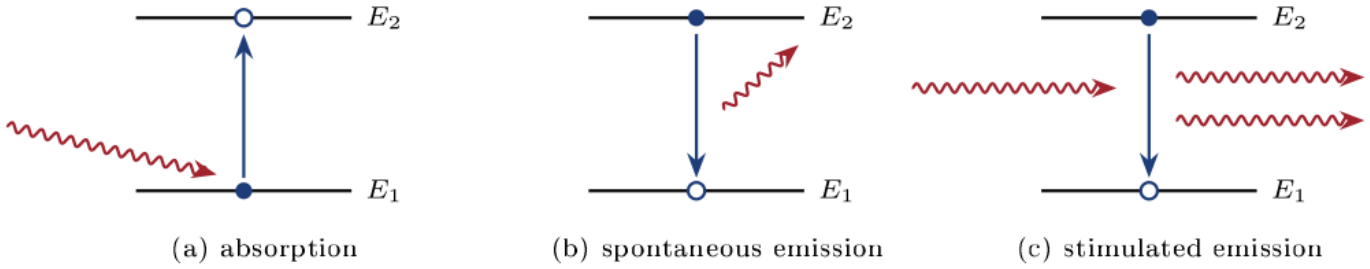


Figure 41.11. The three ways light of energy $hf = E_2 - E_1$ interacts with an atom. (a) Absorption removes a photon. (b) Spontaneous emission releases a photon in a random direction. (c) Stimulated emission: an incoming photon triggers an excited atom to emit a second photon identical to the first, so the light is amplified.

•Figure description

Three side-by-side panels, each showing a lower energy level E sub 1 and an upper level E sub 2 as horizontal lines, with an atom's electron drawn as a dot. Panel a, absorption: a wavy incoming photon arrow strikes an electron on the lower level, and a vertical arrow shows the electron jumping up to E sub 2; no photon leaves. Panel b, spontaneous emission: an electron on the upper level drops to the lower level on its own, and one wavy photon leaves in an arbitrary direction, angled upward. Panel c, stimulated emission: an incoming wavy photon passes an electron on the upper level, which drops to the lower level, and two identical wavy photons leave side by side in the same direction as the incoming one.

Population inversion. In thermal equilibrium at temperature T, the numbers of atoms in two levels separated by $\Delta E = E_2 - E_1$ are in the ratio given by the **Boltzmann factor**:

Key Equation: Boltzmann population ratio

$$\frac{N_2}{N_1} = e^{-\Delta E/k_B T}. \tag{41.21}$$

The upper level is always the less populated one, so in ordinary matter absorption always wins and light is weakened as it passes through. To build a laser we must create the opposite, unnatural situation, a **population inversion** with $N_2 > N_1$, so that stimulated emission outpaces absorption.

Example 41.7: How empty is the upper level?

The red helium–neon laser line comes from a transition with $\Delta E = 1.96\ \text{eV}$. Find N_2/N_1 for these two levels in thermal equilibrium (a) at room temperature, $300\ \text{K}$, and (b) at $3000\ \text{K}$, the temperature of a lamp filament. Use $k_B = 8.617 \times 10^{-5}\ \text{eV/K}$.

Solution

(a) At $300\ \text{K}$, $k_BT = (8.617 \times 10^{-5})(300) = 0.0259\ \text{eV}$, so

$$\frac{N_2}{N_1} = e^{-1.96/0.0259} = e^{-75.8} = \boxed{1 \times 10^{-33}}. \tag{41.7.1}$$

In a whole mole of gas, about 6×10^{23} atoms, not even one would be in the upper level.

(b) At $3000\ \text{K}$, $k_BT = 0.259\ \text{eV}$, and

$$\frac{N_2}{N_1} = e^{-1.96/0.259} = e^{-7.58} = \boxed{5 \times 10^{-4}}. \tag{41.7.2}$$

Much better, but still nowhere near an inversion. Because $e^{-\Delta E/k_BT}$ is less than 1 for every positive temperature, no amount of heating can produce $N_2 > N_1$. A population inversion requires deliberately driving the system away from thermal equilibrium.

The atoms must be **pumped** into the upper level by an outside energy source: an electric discharge, a flash lamp, another laser, or an electric current. Pumping alone is not enough, though. If the upper level emptied by spontaneous emission within nanoseconds, the inversion would vanish as fast as it was built. The trick is to use a **metastable state**, an excited level whose spontaneous decay is unusually slow (lifetimes of microseconds to milliseconds, a million times longer than usual) because the transition is nearly forbidden by the quantum-number rules. Atoms pile up in the metastable level, and an inversion forms.

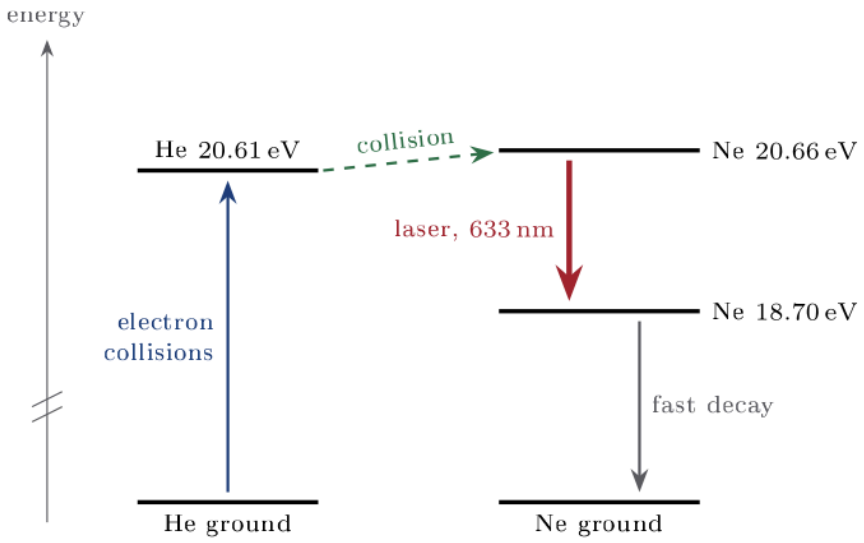


Figure 41.12. The helium–neon laser. Electron collisions in a gas discharge pump helium atoms into a metastable level at $20.61\ \text{eV}$. Collisions hand that energy to neon atoms, whose upper level at $20.66\ \text{eV}$ nearly matches it, producing a population inversion relative to the neon level at $18.70\ \text{eV}$. The $1.96\ \text{eV}$ transition between them emits red light at $633\ \text{nm}$. (The energy axis is broken between 0 and about $18\ \text{eV}$.)

•Figure description

An energy-level diagram for the helium-neon laser, with energy in electron-volts increasing upward. On the left, helium has its ground state at 0 and a metastable level at 20.61 eV; an upward arrow labeled electron collisions in the discharge pumps helium from the ground state to this level. On the right, neon has its ground state at 0, an upper laser level at 20.66 eV, just above the helium level, and a lower laser level at 18.70 eV. A dashed horizontal arrow labeled collision carries energy from the helium metastable level across to the neon upper level. A bold downward red arrow from 20.66 to 18.70 eV is labeled laser, 633 nm. A shorter gray arrow labeled fast decay drops from the lower laser level toward neon's ground state.

The helium–neon laser of Figure 41.12 shows how this works in practice. A glass tube holds a mixture of mostly helium with a little neon, and a high voltage drives an electric discharge through it. Electrons in the discharge knock helium atoms into a metastable level $20.61\ \text{eV}$ above the ground state. Neon happens to have a level at $20.66\ \text{eV}$, almost exactly the same energy, so when an excited helium atom collides with a neon atom, the energy hops across (the small difference comes from the atoms' kinetic energy). Neon atoms therefore accumulate in the $20.66\ \text{eV}$ level. The level below it, at $18.70\ \text{eV}$, empties quickly, so it stays nearly vacant and the two neon levels are inverted. The transition between them releases

$$\lambda = \frac{hc}{\Delta E} = \frac{1240\,\text{eV}\cdot\text{nm}}{20.66 - 18.70\,\text{eV}} = \frac{1240}{1.96}\,\text{nm} = 633\,\text{nm}, \tag{41.7.3}$$

the familiar red of a laboratory He–Ne laser.

The optical cavity. A population inversion makes a medium that amplifies light, but the amplification in a single pass is small. So the medium is placed between two parallel mirrors, forming an **optical cavity**. A spontaneously emitted photon that happens to travel along the axis stimulates a copy, the two stimulate two more, and the growing pulse is reflected back and forth through the medium hundreds of times, building up an intense standing wave (Chapter 30). Photons heading off-axis simply leave through the sides. One mirror transmits a percent or so of the light, and that fraction emerges as the beam.

The result has properties no ordinary source can match.

- *Monochromatic*: every photon comes from the same transition, and the cavity selects a very narrow band of wavelengths within it.
- *Coherent*: the photons are copies of one another, in phase, so the beam can produce the sharp interference patterns of Chapter 33 over large path differences.
- *Directional*: only light traveling along the cavity axis is amplified, so the beam spreads only by diffraction (Chapter 34), a fraction of a milliradian for a typical laser.
- *Intense*: because the power is concentrated in a narrow beam, even a milliwatt laser delivers an intensity at the eye greater than direct sunlight.

Lasers now read barcodes and optical discs, carry the internet's data through optical fibers, cut steel, reshape the cornea, and measure the distance to the Moon to the centimeter by timing pulses reflected from mirrors the Apollo astronauts left behind. Most of the lasers in use today are not gas tubes but tiny semiconductor diode lasers, which we meet in Chapter 42.

Example 41.8: Photons in a laser beam

A He–Ne laser emits $5.0\ \text{mW}$ of $633\ \text{nm}$ light in a beam $1.0\ \text{mm}$ in diameter. (a) How many photons per second does it emit? (b) What is the beam's intensity, and how does it compare with sunlight ($1361\ \text{W/m}^2$)?

Solution

(a) Each photon carries

$$E = \frac{hc}{\lambda} = \frac{1240\,\text{eV}\cdot\text{nm}}{633\,\text{nm}} = 1.96\,\text{eV} = (1.96)(1.602\times10^{-19}\,\text{J}) = 3.14\times10^{-19}\,\text{J}. \tag{41.7.4}$$

The number emitted per second is the power divided by the energy per photon:

$$\frac{N}{t} = \frac{P}{E} = \frac{5.0\times10^{-3}\,\text{J/s}}{3.14\times10^{-19}\,\text{J}} = \boxed{1.6\times10^{16}\,\text{photons/s}}. \tag{41.7.5}$$

(b) The beam's cross-sectional area is $\pi(0.50\times10^{-3}\,\text{m})^2 = 7.9\times10^{-7}\ \text{m}^2$, so

$$I = \frac{P}{A} = \frac{5.0\times10^{-3}}{7.9\times10^{-7}} = \boxed{6.4\times10^{3}\,\text{W/m}^2}, \tag{41.7.6}$$

nearly five times the intensity of full sunlight, from a device that draws a few watts. That is why you must never look into a laser beam: the eye's lens focuses that already-intense, parallel beam onto a tiny spot on the retina.

Checkpoint 41.6

A beam of $633\ \text{nm}$ light passes through a tube of neon gas that is *not* pumped, so nearly every atom is in the lower of the two laser levels. Is the beam amplified or weakened? What changes when the discharge is switched on and the levels become inverted?

Answer

Answers to Checkpoints

Checkpoint 41.6. Weakened. A $633\ \text{nm}$ photon is as likely to be absorbed by an atom in the lower level as to stimulate emission from one in the upper level, so the net effect depends on which level holds more atoms. With nearly all the atoms in the lower level, absorption wins and the beam loses photons. With the discharge on and the levels inverted, stimulated emission outpaces absorption, and each pass through the tube adds photons identical to those already in the beam: the beam is amplified.

This page titled 41.7: Lasers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.S: Atomic Physics (Summary)

Summary

- The model of the atom evolved from Dalton's solid sphere, through Thomson's **plum pudding**, to Rutherford's **nuclear** atom (revealed by the gold-foil experiment), Bohr's quantized orbits, and finally the quantum **electron-cloud** model.
- A bound electron is a trapped particle; its de Broglie wave must form a closed standing wave, $2\pi r = n\lambda$, which quantizes its energy.
- **Hydrogen energy levels:** $E_n = -me^4/8\epsilon_0^2h^2 \cdot 1/n^2 = -13.6\ \text{eV}/n^2$. Levels are negative (bound), with the ground state at $-13.6\ \text{eV}$ and the Bohr radius $a_0 = 52.9\ \text{pm}$.
- A transition emits or absorbs a photon with $\Delta E = E_{\text{high}} - E_{\text{low}} = hf = hc/\lambda$. Wavelengths follow the **Rydberg formula** $1/\lambda = R(1/n_{\text{low}}^2 - 1/n_{\text{high}}^2)$, $R = 1.097 \times 10^7\ \text{m}^{-1}$.
- **Emission** (downward) gives bright lines; **absorption** (upward) gives dark lines at the same wavelengths. The Lyman, Balmer, and Paschen series end on $n = 1, 2, 3$ (UV, visible, IR). Each element's line pattern is a unique **spectral fingerprint**.
- The quantum model replaces orbits with clouds. The ground state $\psi_{1s} = (\pi a_0^3)^{-1/2}e^{-r/a_0}$ has radial probability density $P(r) = (4r^2/a_0^3)e^{-2r/a_0}$, which peaks at $r = a_0$; for it, $\langle r\rangle = 1.5\ a_0$.
- Three quantum numbers arise from separating the wave function: n (energy), $\ell = 0, \ldots, n-1$ (angular-momentum magnitude $L = \sqrt{\ell(\ell+1)}\hbar$), and $m_\ell = -\ell, \ldots, +\ell$ (direction, $L_z = m_\ell\hbar$, giving **space quantization**).
- Electrons also carry **spin** ($m_s = \pm\frac{1}{2}$). The **Pauli exclusion principle** forbids two electrons from sharing all four quantum numbers, so subshells hold $2(2\ell+1)$ electrons and shells hold $2n^2$. This filling builds electron configurations and explains the periodic table.
- A **laser** amplifies light by **stimulated emission**, which copies an incoming photon exactly. In thermal equilibrium $N_2/N_1 = e^{-\Delta E/k_BT} < 1$, so a laser needs a pumped **population inversion**, usually built in a metastable level, inside a mirrored optical cavity. Laser light is monochromatic, coherent, directional, and intense.

Key Equations

Standing wave around a circular orbit

$$2\pi r = n\lambda, \qquad n = 1, 2, 3, \ldots \tag{41.S.1}$$

Bohr quantization of angular momentum

$$L = mvr = n\hbar \tag{41.S.2}$$

Energy levels of hydrogen

$$E_n = -\frac{me^4}{8\epsilon_0^2h^2}\frac{1}{n^2} = \frac{-13.6\,\text{eV}}{n^2} \tag{41.S.3}$$

Bohr radius and orbital radii

$$a_0 = \frac{\epsilon_0h^2}{\pi me^2} = 52.9\,\text{pm}, \qquad r_n = n^2a_0 \tag{41.S.4}$$

Photon emitted or absorbed in a transition

$$\Delta E = E_{\text{high}} - E_{\text{low}} = hf = \frac{hc}{\lambda} \tag{41.S.5}$$

The Rydberg formula

$$\frac{1}{\lambda} = R\left(\frac{1}{n_{\text{low}}^2} - \frac{1}{n_{\text{high}}^2}\right), \qquad R = 1.097373 \times 10^7\,\text{m}^{-1} \tag{41.S.6}$$

Hydrogen ground state and its radial probability density

$$\begin{aligned} \psi_{1s}(r) &= \frac{1}{\sqrt{\pi}\,a_0^{3/2}}\,e^{-r/a_0} \\ P(r) &= \frac{4r^2}{a_0^3}\,e^{-2r/a_0} \qquad (\text{peak at } r = a_0,\ \langle r\rangle = 1.5\,a_0) \end{aligned} \tag{41.S.7}$$

Allowed quantum numbers

$$\begin{aligned} n &= 1, 2, 3, \ldots \\ \ell &= 0, 1, 2, \ldots, n-1 \\ m_\ell &= -\ell, -\ell+1, \ldots, +\ell \\ m_s &= \pm\tfrac{1}{2} \end{aligned} \tag{41.S.8}$$

Magnitude and direction of orbital angular momentum

$$L = \sqrt{\ell(\ell+1)}\,\hbar, \qquad L_z = m_\ell\,\hbar, \qquad \cos\theta = \frac{m_\ell}{\sqrt{\ell(\ell+1)}} \tag{41.S.9}$$

Subshell and shell capacities

$$2(2\ell+1)\ \text{per subshell}, \qquad 2\sum_{\ell=0}^{n-1}(2\ell+1) = 2n^2\ \text{per shell} \tag{41.S.10}$$

Boltzmann population ratio

$$\frac{N_2}{N_1} = e^{-\Delta E/k_B T} \tag{41.S.11}$$

This page titled 41.S: Atomic Physics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

41.E: Atomic Physics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $h = 6.63 \times 10^{-34}\ \text{J}\cdot\text{s}$, $\hbar = 1.055 \times 10^{-34}\ \text{J}\cdot\text{s}$, $m_e = 9.11 \times 10^{-31}\ \text{kg}$, $e = 1.602 \times 10^{-19}\ \text{C}$, $c = 3.00 \times 10^{8}\ \text{m/s}$, $a_0 = 5.29 \times 10^{-11}\ \text{m}$, $R = 1.097 \times 10^{7}\ \text{m}^{-1}$, $1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J}$, $hc = 1240\ \text{eV}\cdot\text{nm}$, and $k_B = 8.617 \times 10^{-5}\ \text{eV/K}$.

Models of the Atom

41.1 ★ *[Conceptual.]* List, in order, the five models of the atom discussed in this chapter, and state in one phrase the key new feature each one introduced.

41.2 ★★ *[Conceptual.]* In Rutherford's gold-foil experiment, roughly 1 alpha particle in 8000 was deflected by more than 90°. Explain how this small fraction, rather than a large one, is what reveals that the nucleus is extremely small compared with the atom.

Energies of a Trapped Electron and the Hydrogen Atom

41.3 ★ Find the energy (in eV) of the $n = 1$, $n = 2$, and $n = 3$ levels of hydrogen.

41.4 ★★ An electron orbits a hydrogen nucleus in the $n = 4$ state. Using the standing-wave condition $2\pi r = n\lambda$ with $r = n^2 a_0$, find the de Broglie wavelength of the electron.

41.5 ★★ Find the ionization energy of hydrogen: the energy needed to remove the electron from the ground state to $n = \infty$.

41.6 ★★ The electron in a hydrogen atom drops from $n = 4$ to $n = 2$. Find the energy, wavelength, and color region of the emitted photon.

41.7 ★★★ Show, starting from the centripetal condition $mv^2/r = e^2/4\pi\epsilon_0 r^2$ and the quantization $mvr = n\hbar$, that the orbital radius is $r_n = n^2 a_0$ with $a_0 = \epsilon_0 h^2/\pi m e^2$.

Energy Level Transitions and Atomic Spectra

41.8 ★ *[Conceptual.]* A cool hydrogen gas is placed between a white-light source and a spectrometer. Will the spectrum show bright lines, dark lines, or both? Explain.

41.9 ★★ *[Real-world application: gas-discharge lamps.]* In a discharge tube, ground-state hydrogen atoms are struck by electrons that each carry $12.5\ \text{eV}$ of kinetic energy. Unlike a photon, a colliding electron can give up just part of its energy, so an atom can be raised to any level whose energy above the ground state is less than $12.5\ \text{eV}$.

a. What is the highest level n the atoms can reach?
b. List every photon wavelength the gas can then emit as the atoms return to the ground state.
c. Which of these lines are visible?
d. What is the smallest electron energy that would add a second visible line, and what is its wavelength?

41.10 ★★ The Balmer series limit is the shortest-wavelength line in the series, emitted in the transition $n = \infty \to 2$. Find its wavelength.

41.11 ★★★ *[Real-world application: astronomy.]* A particular Balmer line of hydrogen, normally at $486.1\ \text{nm}$, is observed from a distant galaxy at $511.0\ \text{nm}$. Treating the shift as a low-speed Doppler effect ($\Delta\lambda/\lambda \approx v/c$), estimate the galaxy's recession speed.

The Quantum Model and Wave Functions of Hydrogen

41.12 ★ Write the four quantum numbers and state the allowed values of each.

41.13 ★★ How many distinct (n, ℓ, m_ℓ) orbital states are there in the $n = 4$ shell? Verify that your count equals n^2.

41.14 ★★ List all the allowed subshells (e.g. $2s$, $2p$) for $n = 3$, and give the ℓ value of each.

41.15 ★★★ Show by direct integration that the hydrogen ground-state wave function $\psi_{1s}(r) = (\pi a_0^3)^{-1/2} e^{-r/a_0}$ is normalized, i.e. that $\int_0^\infty |\psi_{1s}|^2 \, 4\pi r^2 \, dr = 1$. (You may use $\int_0^\infty u^2 e^{-u} \, du = 2$.)

Physical Interpretation of Quantum Numbers

41.16 ★★ A hydrogen atom is in an $\ell = 2$ state. Find the magnitude of its orbital angular momentum and the allowed values of L_z.

41.17 ★★ *[The classical limit.]* The smallest angle $\theta_{\min}$ between $\mathbf{L}$ and the z-axis occurs when $m_\ell = \ell$.

a. Find $\theta_{\min}$ for $\ell = 1$, $\ell = 10$, and $\ell = 100$.
b. Show that $\tan\theta_{\min} = 1/\sqrt{\ell}$ exactly, so that for large ℓ, $\theta_{\min} \approx 1/\sqrt{\ell}$ (in radians). Check this against your $\ell = 100$ result.
c. Explain how your results illustrate the correspondence principle.

41.18 ★★★ *[Conceptual.]* Explain why the orbital angular-momentum vector $\mathbf{L}$ can never point exactly along the z-axis, and connect your answer to the uncertainty principle.

The Pauli Exclusion Principle

41.19 ★ How many electrons can occupy (a) a d subshell, (b) the $n = 2$ shell, (c) the $n = 4$ shell?

41.20 ★★ Write the ground-state electron configuration of oxygen ($Z = 8$) and identify the number of valence electrons.

41.21 ★★ *[Conceptual.]* Explain, using the Pauli exclusion principle, why the second row of the periodic table contains exactly eight elements.

Lasers

41.22 ★ *[Conceptual.]* Name three ways the light of a laser differs from the light of an incandescent bulb, and for each, identify the feature of the laser (stimulated emission, the population inversion, or the optical cavity) that is responsible.

41.23 ★★ A green laser pointer emits $1.0\ \text{mW}$ at $532\ \text{nm}$. (a) Find the energy of each photon in eV and in joules. (b) How many photons does it emit per second? (c) A red pointer of the same power emits at $650\ \text{nm}$. Does it emit more photons per second or fewer, and by what factor?

41.24 ★★ The ruby laser, the first laser ever built (1960), emits at $694.3\ \text{nm}$. Use $k_B = 8.617 \times 10^{-5}\ \text{eV/K}$. (a) Find the energy difference between the two laser levels. (b) Find the ratio N_2/N_1 for these levels in thermal equilibrium at $300\ \text{K}$. (c) At what temperature would the upper level hold half as many atoms as the lower one? (d) Explain why no temperature, however high, can produce a population inversion.

Spectral Fingerprints — the Chapter Callback

41.25 ★★★ *[Real-world application: spectroscopy — the chapter callback.]* An astronomer records four emission lines from a glowing cloud of gas, at wavelengths 410, 434, 486, and $656\ \text{nm}$.

a. Show that each of these wavelengths corresponds to a Balmer transition of hydrogen ($n_{\text{high}} \to 2$), and identify the value of n_{high} for each line.
b. Explain how the astronomer can be confident the cloud contains hydrogen, even without sampling it, and relate this to how helium was first discovered in the Sun.
c. The cloud also shows a strong line at $587.6\ \text{nm}$ that fits no hydrogen series. What can the astronomer conclude, and what would be the next step to identify its source?

Challenge Problems

41.26 ★★★ *[Real-world application: fusion research — hydrogen-like ions.]* An ion stripped of all but one electron is still a two-body Coulomb problem, so the Bohr results carry over with the nuclear charge e replaced by Ze.

a. Repeat the Bohr argument with a nucleus of charge $+Ze$ to show that

$$E_n = -\frac{(13.6\,\text{eV})\,Z^2}{n^2}, \qquad r_n = \frac{n^2 a_0}{Z}. \tag{41.E.1}$$

b. Find the energy and wavelength of the $n = 2 \to 1$ (Lyman-α) photon of singly ionized helium, He^+ ($Z = 2$). In what part of the spectrum does it lie?
c. A tokamak plasma is contaminated by iron sputtered from the vessel wall, some of it stripped to hydrogen-like Fe^{25+} ($Z = 26$). Find the energy and wavelength of its Lyman-α photon, and its ground-state orbital radius.
d. Fusion reactors are engineered obsessively to keep high-Z impurities out of the plasma. Using your answers, explain why even a trace of iron is far more damaging to the energy balance than the same number of stray hydrogen atoms would be.

41.27 ★★★ *[Real-world application: astronomy — giant atoms in interstellar space.]* In the tenuous gas of an ionized nebula, a proton occasionally captures an electron into a very high Bohr level, which then cascades downward. Radio astronomers observe the resulting *recombination lines*. Consider the transition $n = 110 \to n = 109$ in hydrogen.

a. Find the orbital radius of the $n = 110$ state, in micrometers, and compare it with the size of a typical bacterium ($\sim 1\ \mu\text{m}$). By what factor is this atom larger than a ground-state hydrogen atom?
b. Find the energy of the emitted photon in eV. *(Keep enough digits: you are subtracting two nearly equal numbers.)*
c. Find the frequency and wavelength of the photon. Confirm that it lies in the radio band.
d. Such atoms are never seen in a laboratory gas at ordinary pressure. Estimate the mean spacing between particles that would be needed for an atom this size to survive long enough to radiate, and explain what this tells you about conditions in the nebula.

41.28 ★★★ *[Real-world application: materials — where magnetism comes from.]* The Pauli exclusion principle, together with the order in which subshells fill, determines the entire structure of the periodic table and, as it happens, which metals are magnetic.

a. Using the allowed values of ℓ and m_ℓ, show that a shell of principal quantum number n holds $2n^2$ electrons, and evaluate this for $n = 1, 2, 3, 4$.
b. If shells simply filled in order of n, the third row of the periodic table would contain 18 elements. It contains 8. Explain what actually happens, referring to the relative energies of the $3d$ and $4s$ subshells.
c. Write the ground-state electron configuration of iron ($Z = 26$). Using Hund's rule (that electrons occupy the orbitals of a subshell singly, with parallel spins, before pairing up) determine how many unpaired electrons the $3d$ subshell contains.
d. Zinc ($Z = 30$) is not magnetic, though it sits only four places after iron. Write its configuration, count its unpaired electrons, and explain the difference. What does your answer suggest about where in the periodic table to look for magnetic elements?

This page titled 41.E: Atomic Physics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42: BONDING AND ELECTRICAL CONDUCTION IN SOLIDS

A diamond and the “lead” of a pencil are made of exactly the same thing: pure carbon. Yet pass a current through them and they could hardly be more different. Graphite, the pencil lead, conducts electricity well enough to be used for the brushes of electric motors. Diamond is one of the best electrical insulators known; its resistivity is more than a hundred billion billion times larger, a difference of about 21 orders of magnitude. Copper, a metal, conducts better still, and silicon, the stuff of computer chips, sits awkwardly in between: a poor conductor whose conductivity can be raised a millionfold by adding one impurity atom for every ten million silicon atoms. No other property of ordinary materials ranges so widely. The difference between diamond and graphite cannot lie in the atoms, which are identical. It lies in how the atoms are bonded together, and in what that bonding does to the energies their electrons are allowed to have.

This chapter applies the quantum mechanics of the last three chapters to matter in bulk. We begin with how atoms bind into molecules, and why molecules vibrate only in quantized steps (which resolves a puzzle about the heat capacity of gases from Chapter 36). We then treat the electrons of a metal as particles in a box, filled according to the Pauli exclusion principle, and find that they move at astonishing speeds even at absolute zero. Bringing many atoms together turns their discrete energy levels into **bands**, and whether a band is full or partly full decides whether a solid is a metal, an insulator, or a semiconductor. The chapter ends with the semiconductor devices (the diode, the light-emitting diode, the transistor, and the integrated circuit) that have reshaped the world.

Learning Objectives

- Describe ionic, covalent, van der Waals, hydrogen, and metallic bonds in terms of a potential-energy curve, and estimate an ionic bond energy.
- Calculate the quantized vibrational energies of a molecule and explain why vibration is frozen out at room temperature.
- Find the Fermi energy and Fermi speed of the free electrons in a metal, and use the Fermi–Dirac distribution.
- Explain how energy bands form, and use the band structure to classify a solid as a conductor, insulator, or semiconductor.
- Describe intrinsic and doped semiconductors, and explain how a p–n junction diode, a light-emitting diode, and a transistor work.

We will return to diamond, graphite, copper, and silicon once we have the band picture, and again in the final problems.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

42.1: Molecular Bonds
42.2: Free-Electron Theory of Metals
42.3: Electrical Conduction in Metals and Insulators
42.4: Semiconductors
42.S: Bonding and Electrical Conduction in Solids (Summary)
42.E: Bonding and Electrical Conduction in Solids (Problems)

This page titled 42: Bonding and Electrical Conduction in Solids is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.1: Molecular Bonds

Two atoms form a **bond** when the pair has less total energy close together than far apart. Whatever the kind of bond, its energy as a function of the separation r between the nuclei has the same general shape (Figure 42.1). At large separations the potential energy $U(r)$ is zero; the atoms ignore each other. As they approach, an attractive force pulls them together and U falls. At very close range, a strong repulsion takes over and U climbs steeply. Part of the repulsion is the Coulomb push between the two nuclei, but most of it comes from the Pauli exclusion principle (Chapter 41): when the filled inner shells of the two atoms begin to overlap, their electrons cannot share the same states, so some are forced into higher-energy states. The balance point, the minimum of the curve, is the **equilibrium separation** r_0, and the depth of the well below zero is the energy needed to pull the atoms apart, the **dissociation energy** (or bond energy) of the molecule.

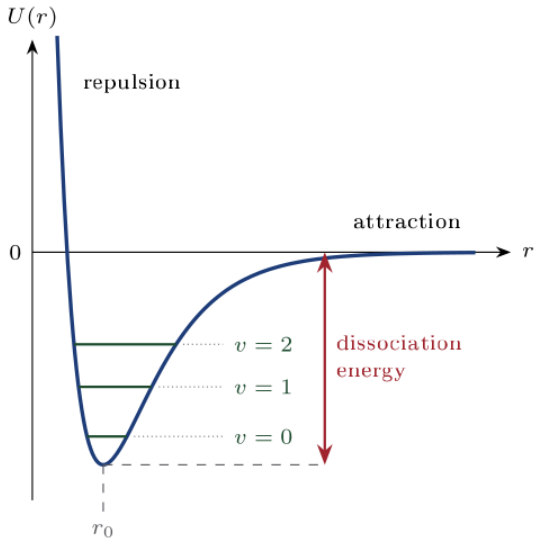


Figure 42.1. The potential energy of two atoms as a function of their separation. Attraction pulls them together until short-range repulsion, mostly from the Pauli exclusion principle, stops them at the equilibrium separation r_0. The depth of the well is the dissociation energy. A molecule can vibrate only with the quantized energies marked $v = 0, 1, 2, ...$; even the lowest lies above the bottom of the well.

•Figure description

A graph of potential energy U against the separation r between two nuclei. The curve starts very high at small r, plunges steeply to a minimum below zero at the equilibrium separation r sub 0, and then rises gradually, leveling off toward zero at large r. The steep left wall is labeled repulsion and the gentle right side is labeled attraction. A vertical double arrow from the minimum up to the zero line is labeled dissociation energy. Inside the well, three short horizontal lines mark the lowest allowed vibrational energies, v equals 0, 1 and 2, each spanning the width of the well at its height; their spacing shrinks slightly going up, and the lowest one sits above the bottom of the well.

What supplies the attraction depends on the atoms, and it is the basis for classifying bonds.

Ionic bonds. An **ionic bond** forms when one atom gives up an electron to another and the resulting positive and negative ions attract each other electrically. Sodium chloride is the classic case. Sodium has a single valence electron outside a closed shell (Chapter 41), which takes only $5.14\ \text{eV}$ to remove. Chlorine is one electron short of a closed shell, and it *releases* $3.61\ \text{eV}$, its **electron affinity**, when it captures an extra electron. Transferring an electron from Na to Cl therefore costs a net $5.14 - 3.61 = 1.53\ \text{eV}$. That cost is repaid many times over by the Coulomb attraction of the ions, $U = -ke^2/r$, once they are close together.

For atomic-scale calculations it helps to know that

$$ke^2 = 1.44\,\text{eV} \cdot \text{nm}, \tag{42.1.1}$$

so the electrical potential energy of two singly charged ions at a separation r in nanometers is $-1.44/r$ in electron-volts.

Example 42.1: The bond in a sodium chloride molecule

In a molecule of NaCl the ions are $0.236\ \text{nm}$ apart. (a) Estimate the energy released in forming the molecule from neutral Na and Cl atoms, ignoring the short-range repulsion. (b) At what separation does transferring the electron first become energetically favorable?

Solution

(a) The Coulomb energy of the ion pair at r_0 is

$$U = -\frac{ke^2}{r_0} = -\frac{1.44\,\text{eV}\cdot\text{nm}}{0.236\,\text{nm}} = -6.10\,\text{eV}. \tag{42.1.2}$$

Making the ions in the first place cost $1.53\ \text{eV}$, so the energy released is about

$$6.10 - 1.53 = \boxed{4.6\,\text{eV}}. \tag{42.1.3}$$

The measured dissociation energy is $4.26\ \text{eV}$. The difference of about $0.3\ \text{eV}$ is the repulsion of the overlapping closed shells, which we ignored.

(b) The transfer pays off once the Coulomb energy exceeds the $1.53\ \text{eV}$ cost:

$$r < \frac{1.44\,\text{eV}\cdot\text{nm}}{1.53\,\text{eV}} = \boxed{0.94\,\text{nm}}, \tag{42.1.4}$$

about four times the final bond length. Inside that distance the electron effectively jumps from the sodium to the chlorine, and the ions then pull each other in to r_0.

Covalent bonds. Two identical atoms, such as the two hydrogen atoms of H_2, have no reason to trade an electron. Instead they *share* a pair. In a **covalent bond** the wave functions of the two atoms' electrons combine into a molecular state concentrated between the nuclei, where each electron is attracted by both nuclei at once. The Pauli exclusion principle allows two electrons to occupy that one spatial state only if their spins are opposite, which is why covalent bonds are built from electron *pairs*. The H_2 bond has $r_0 = 0.074\ \text{nm}$ and a dissociation energy of $4.52\ \text{eV}$. Carbon, with four valence electrons, can form four covalent bonds pointing toward the corners of a tetrahedron; in diamond every carbon atom is bonded this way to four neighbors, locking every valence electron into a bond. Covalent bonds are strong and directional, and covalent solids such as diamond and silicon are hard and have high melting points.

Weaker bonds. Even atoms with closed shells, such as argon, attract each other weakly. The electron cloud of each atom fluctuates, creating a momentary electric dipole that induces a dipole in its neighbor, and the two dipoles attract. These **van der Waals bonds** are a hundred times weaker than covalent ones, which is why argon becomes a liquid only below $87\ \text{K}$. Molecules with permanent dipoles attract more strongly. In a **hydrogen bond**, a hydrogen atom covalently bonded to a strongly electronegative atom (oxygen, nitrogen, or fluorine) is left as a nearly bare proton on one side, and it attracts a lone electron pair on a neighboring molecule. Hydrogen bonds hold water molecules together into the open structure of ice (Chapter 36) and hold together the two strands of DNA, strongly enough to be stable and weakly enough to be unzipped when a cell divides.

Metallic bonds. In a metal, each atom gives up one or more valence electrons, not to a particular neighbor but to the crystal as a whole. The positive ions sit in a regular lattice immersed in a "sea" of electrons that belong to no atom in particular, and the attraction between the ions and the electron sea holds the metal together. This **metallic bond** is not directional, which is why metals can be bent and hammered into shape without shattering, and its free electrons are what make metals conduct. Section 42.2 takes them up.

Table 42.1. Types of bond, with typical bond energies per bond (or per ion pair).

Bond	Example	Typical energy	Origin of the attraction
Ionic	NaCl	\SIrange38eV	Coulomb attraction of ions
Covalent	H_2, diamond	\SIrange210eV	shared electron pair
Metallic	Cu, Na	\SIrange15eV	ions in a sea of electrons

Bond	Example	Typical energy	Origin of the attraction
Hydrogen	water, DNA	$\SIrange{0.1}{0.4}{eV}$	nearly bare proton and lone pair
Van der Waals	solid Ar	$\SIrange{0.01}{0.1}{eV}$	induced fluctuating dipoles

Checkpoint 42.1

Solid sodium chloride melts at 801°C, while solid argon melts at -189°C. Use Table 42.1 to explain the enormous difference.

Answer

Molecular vibrations. Near the bottom of the well in Figure 42.1, the potential-energy curve is very nearly a parabola, $U \approx U_{\min} + \frac{1}{2}k(r-r_0)^2$. The bond acts like a spring of stiffness k, and the two atoms oscillate along it as a simple harmonic oscillator (Chapter 28) with frequency

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}, \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}, \tag{42.1}$$

where μ, the **reduced mass**, accounts for both atoms moving. (For two identical atoms of mass m, $\mu = m/2$.) But a molecule is a quantum system, and like the electron in a well (Chapter 40), its vibrational energy is quantized. Solving Schrödinger's equation for a parabolic potential gives equally spaced levels:

Key Equation: Vibrational energy levels

$$E_v = \left(v + \tfrac{1}{2}\right) hf, \qquad v = 0, 1, 2, \ldots \tag{42.2}$$

Two features stand out. The levels are spaced by exactly hf, so a molecule can absorb or emit vibrational energy only in quanta of hf; these transitions fall in the infrared, and they are why carbon dioxide and water vapor absorb the infrared radiation from Earth's surface. And the lowest level, $v = 0$, has energy $\frac{1}{2}hf$, not zero: like any confined particle, the molecule can never be perfectly still. (Higher up, where the real well widens, the levels crowd slightly closer together, as Figure 42.1 shows.)

Example 42.2: Why nitrogen does not vibrate

The nitrogen molecule N_2 vibrates at $f = 7.07 \times 10^{13}\ \text{Hz}$. (a) Find the spacing of its vibrational levels in eV. (b) Find the stiffness of the bond. Each nitrogen atom has mass $2.33 \times 10^{-26}\ \text{kg}$. (c) In a gas at $300\ \text{K}$, what fraction of the molecules are in the first excited vibrational level? Use the Boltzmann factor $e^{-\Delta E/k_B T}$ (Chapter 41).

Solution

(a)

$$hf = (6.626 \times 10^{-34}\,\text{J}\cdot\text{s})(7.07 \times 10^{13}\,\text{Hz}) = 4.68 \times 10^{-20}\,\text{J} = \boxed{0.292\,\text{eV}}. \tag{42.1.5}$$

(b) For two identical atoms $\mu = m/2 = 1.165 \times 10^{-26}\ \text{kg}$. Solving equation (42.1) for k,

$$k = (2\pi f)^2 \mu = (4.44 \times 10^{14}\,\text{s}^{-1})^2 (1.165 \times 10^{-26}\,\text{kg}) = \boxed{2.3 \times 10^{3}\,\text{N/m}}, \tag{42.1.6}$$

about as stiff as a car's suspension spring, for a "spring" a tenth of a nanometer long.

(c) At $300\ \text{K}$, $k_B T = 0.0259\ \text{eV}$, so the step is $0.292/0.0259 = 11.3$ times $k_B T$, and

$$\frac{N_1}{N_0} = e^{-11.3} \approx \boxed{1 \times 10^{-5}}. \tag{42.1.7}$$

Only about one molecule in a hundred thousand is vibrating. Collisions in a room-temperature gas almost never carry enough energy to excite the vibration, so it contributes nothing to the heat capacity. This is the “frozen-out” degree of freedom of Chapter 36, and it explains why the measured C_V of nitrogen is $\frac{5}{2}R$ rather than $\frac{7}{2}R$. Only above about $1000\ \mathrm{K}$ does the vibration begin to wake up.

Answers to Checkpoints

Checkpoint 42.1. Melting requires enough thermal energy to break the bonds that hold particles in place. Sodium chloride is held together by ionic bonds of several electron-volts per ion pair, while solid argon is held only by van der Waals bonds of about a hundredth of an electron-volt, hundreds of times weaker. Argon’s bonds are broken by thermal agitation at very low temperatures; sodium chloride needs to be red-hot.

This page titled 42.1: Molecular Bonds is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.2: Free-Electron Theory of Metals

The simplest model of a metal takes the metallic bond at face value. Each atom contributes its valence electrons to a common pool, and those electrons move freely throughout the solid, confined only by its surfaces. The metal is then a three-dimensional box full of electrons: exactly the three-dimensional infinite well of Chapter 40. This **free-electron model** ignores the forces of the ions and of the other electrons, which seems a drastic simplification, but it works surprisingly well.

For a cube of side L, the allowed energies are

$$E = \frac{h^2}{8mL^2}\left(n_x^2 + n_y^2 + n_z^2\right), \tag{42.2.1}$$

with each quantum number a positive integer. A piece of metal holds an enormous number of electrons, about 10^{23} in a cubic centimeter of copper, and the Pauli exclusion principle allows only two of them (one of each spin) in each state (n_x, n_y, n_z). So at absolute zero the electrons cannot all sit in the ground state. They fill the states from the bottom up, two per state, until every electron is placed. The energy of the highest filled state at $T = 0$ is the **Fermi energy** E_F.

To find it, picture each state as a point (n_x, n_y, n_z) in a three-dimensional "number space." The states with energy less than E are the points within a distance $R = \sqrt{8mL^2E/h^2}$ of the origin, in the octant where all three numbers are positive. There is one point per unit volume of number space, and two electrons per point, so the number of electrons with energy below E is

$$N = 2 \times \frac{1}{8} \times \frac{4}{3}\pi R^3 = \frac{\pi}{3}\left(\frac{8mL^2E}{h^2}\right)^{3/2}. \tag{42.2.2}$$

Setting $E = E_F$, dividing by the volume $V = L^3$, and solving for E_F gives a result that depends only on the number of free electrons per unit volume, $n = N/V$:

Key Equation: Fermi energy of a metal

$$E_F = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}. \tag{42.3}$$

The size of the box has dropped out, as it should: a large piece of copper and a small piece have the same Fermi energy.

Example 42.3: The electrons in copper

Copper has one free electron per atom, so $n = 8.49 \times 10^{28}\ \mathrm{m^{-3}}$ (the same value we used for drift speed in Chapter 20). Find (a) the Fermi energy, (b) the speed of an electron with that kinetic energy, the **Fermi speed**, and (c) the temperature at which $k_B T$ would equal E_F, the **Fermi temperature**.

Solution

(a) The two factors are

$$\frac{h^2}{2m} = \frac{(6.626 \times 10^{-34})^2}{2(9.109 \times 10^{-31})} = 2.410 \times 10^{-37}\ \mathrm{J \cdot m^2}, \tag{42.2.3}$$

$$\left(\frac{3n}{8\pi}\right)^{2/3} = \left(1.013 \times 10^{28}\ \mathrm{m^{-3}}\right)^{2/3} = 4.68 \times 10^{18}\ \mathrm{m^{-2}}, \tag{42.2.4}$$

so

$$E_F = (2.410 \times 10^{-37})(4.68 \times 10^{18}) = 1.13 \times 10^{-18}\ \mathrm{J} = \boxed{7.0\ \mathrm{eV}}. \tag{42.2.5}$$

(b) Setting $\frac{1}{2}mv_F^2 = E_F$,

$$v_F = \sqrt{\frac{2E_F}{m}} = \sqrt{\frac{2(1.13\times10^{-18})}{9.109\times10^{-31}}} = \boxed{1.6\times10^{6}\,\mathrm{m/s}}, \tag{42.2.6}$$

about half a percent of the speed of light.

(c)

$$T_F = \frac{E_F}{k_B} = \frac{7.0\,\mathrm{eV}}{8.617\times10^{-5}\,\mathrm{eV/K}} = \boxed{8.2\times10^{4}\,\mathrm{K}}. \tag{42.2.7}$$

These numbers are startling. Even at absolute zero, the most energetic electrons in copper move at $1.6\times10^{6}\ \mathrm{m/s}$: the random speed "of order $10^{6}\ \mathrm{m/s}$" quoted in Chapter 20. A classical gas would need a temperature of $80000\ \mathrm{K}$, hotter than the surface of any ordinary star, for its particles to carry this much energy. The electrons move this fast not because they are hot but because the exclusion principle forces them to stack up into high-energy states.

Checkpoint 42.2

Sodium, like copper, has one free electron per atom, but its atoms are farther apart: $n = 2.65\times10^{28}\ \mathrm{m^{-3}}$. Is sodium's Fermi energy larger or smaller than copper's? By roughly what factor?

Answer

The Fermi–Dirac distribution. At absolute zero every state below E_F is filled and every state above it is empty. At a temperature T, thermal energy can lift some electrons into higher states, but only by amounts of order k_BT, which at room temperature is $0.026\ \mathrm{eV}$. An electron deep below E_F cannot be excited by so little, because every state within reach is already occupied and the exclusion principle forbids it to move there. Only electrons within a few k_BT of the top can find empty states to move into. The probability that a state of energy E is occupied is given by the **Fermi–Dirac distribution**:

Key Equation: Fermi–Dirac occupation probability

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT}+1}. \tag{42.4}$$

Well below E_F the exponential is tiny and $f \approx 1$; well above it the exponential is huge and $f \approx 0$; and at $E = E_F$ exactly, $f = \frac{1}{2}$. As Figure 42.2 shows, a rise in temperature merely rounds off the corner of the $T = 0$ step over a width of a few k_BT.

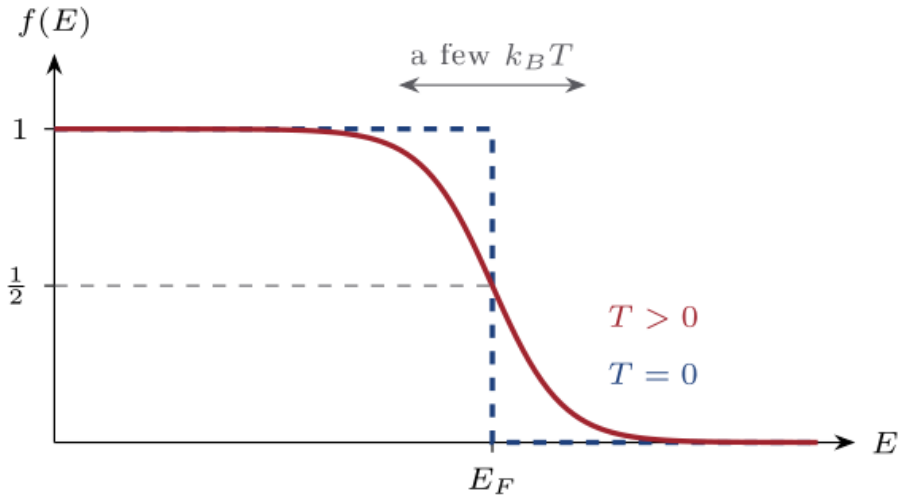


Figure 42.2. The Fermi–Dirac distribution. At $T = 0$ (dashed) every state below the Fermi energy is filled and every state above it is empty. At a higher temperature (solid) the edge is smeared over a few k_BT, but $f(E_F) = \frac{1}{2}$ always. At room temperature k_BT is only about 0.4% of copper's Fermi energy, so on this scale the real smearing would be narrower still.

•Figure description

A graph of the occupation probability f of E against energy E. At absolute zero, drawn as a dashed blue line, f equals 1 for all energies below the Fermi energy E sub F and drops vertically to 0 at E sub F, a sharp step. At a temperature T above zero, drawn as a solid red curve, the step is rounded: f stays near 1 well below E sub F, passes through one half exactly at E sub F, and falls smoothly to near 0 a little above it. A bracket shows that the rounding spans a few k sub B T. Only electrons in this narrow region near E sub F are disturbed by heating.

This resolves a puzzle that troubled physicists for decades. Classically, every free electron in a metal should share in the thermal energy, adding $\frac{3}{2}k_B$ apiece to the heat capacity. Measurements show the electrons contribute almost nothing at room temperature. The Fermi–Dirac distribution explains why: only the small fraction of electrons within about k_BT of E_F, roughly $k_BT/E_F \approx 0.4\%$ in copper, can absorb thermal energy at all. The rest are locked in place by the exclusion principle, just as the vibrations of nitrogen are locked by the size of their quantum step.

Checkpoint 42.3

For a metal at $300\ \text{K}$ ($k_BT = 0.0259\ \text{eV}$), find the probability that a state is occupied (a) at $E = E_F$ and (b) at $E = E_F - 0.20\ \text{eV}$.

Answer

Answers to Checkpoints

Checkpoint 42.2. Smaller. The Fermi energy goes as $n^{2/3}$, so the ratio of sodium's Fermi energy to copper's is $(2.65/8.49)^{2/3} = (0.312)^{2/3} \approx 0.46$. Sodium's Fermi energy is about $0.46 \times 7.0\ \text{eV} \approx 3.2\ \text{eV}$, less than half of copper's. With the electrons spread more thinly, they need not be stacked so high.

Checkpoint 42.3. (a) At $E = E_F$ the exponent is zero, so $f = 1/(1+1) = \frac{1}{2}$, at any temperature. (b) $(E - E_F)/k_BT = -0.20/0.0259 = -7.7$, so $f = 1/(e^{-7.7} + 1) = 1/(1.00045) \approx 0.9995$. A state just $0.2\ \text{eV}$ below the Fermi energy is empty only about 0.05% of the time.

This page titled 42.2: Free-Electron Theory of Metals is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.3: Electrical Conduction in Metals and Insulators

The free-electron model explains why metals conduct, and it even gives a sensible resistivity, but it cannot explain why diamond does not. Every solid is full of electrons. What decides whether they can carry a current?

Resistance in a metal. In Chapter 20 we pictured the conduction electrons of a metal colliding constantly with its ions, each collision erasing the small drift velocity the electric field had given the electron. If τ is the mean free time between collisions, Drude's model gave the resistivity

$$\rho = \frac{m}{ne^2\tau} \tag{42.3.1}$$

(equation (20.7)). We can now ask what that model leaves unexplained.

Example 42.4: How far does an electron go between collisions?

Copper has a resistivity of $1.68 \times 10^{-8}\ \Omega \cdot \mathrm{m}$ at room temperature. Using $n = 8.49 \times 10^{28}\ \mathrm{m}^{-3}$ and the Fermi speed of $1.57 \times 10^{6}\ \mathrm{m/s}$, find (a) the mean time between collisions and (b) the mean free path of a conduction electron. Compare it with the spacing of copper atoms, $0.26\ \mathrm{nm}$.

Solution

(a) Solving $\rho = m/(ne^2\tau)$ for τ,

$$\tau = \frac{m}{ne^2\rho} = \frac{9.109 \times 10^{-31}}{(8.49 \times 10^{28})(1.602 \times 10^{-19})^2(1.68 \times 10^{-8})} = \boxed{2.5 \times 10^{-14}\,\mathrm{s}}. \tag{42.3.2}$$

(b) The electrons that carry current are those near the Fermi energy, moving at v_F, so

$$\lambda = v_F\tau = (1.57 \times 10^{6})(2.5 \times 10^{-14}) = 3.9 \times 10^{-8}\,\mathrm{m} = \boxed{39\,\mathrm{nm}}, \tag{42.3.3}$$

about 150 atomic spacings. A classical electron weaving between the ions would hit one almost every spacing. The quantum explanation is that an electron is a wave, and a wave passes through a *perfectly regular* lattice without scattering at all, just as light passes undeflected through a perfect crystal. Electrons scatter only from what spoils the regularity: impurities, defects, and the thermal vibrations of the ions. Since the vibrations grow with temperature, the resistivity of a metal rises as it warms.

Energy bands. To see why some solids do not conduct at all, we need to take the ions seriously. Bring two identical atoms together, and each of their atomic energy levels splits into two slightly different levels, because the electron's wave function can combine across the two atoms in two ways. Bring together six atoms, and each level splits into six. In a solid of N atoms, with N around 10^{23}, each atomic level splits into N levels so closely spaced that they form an effectively continuous **energy band** (Figure 42.3). The outer levels, whose wave functions overlap most, spread into the widest bands. Between the bands lie **band gaps**, ranges of energy in which no electron states exist at all.

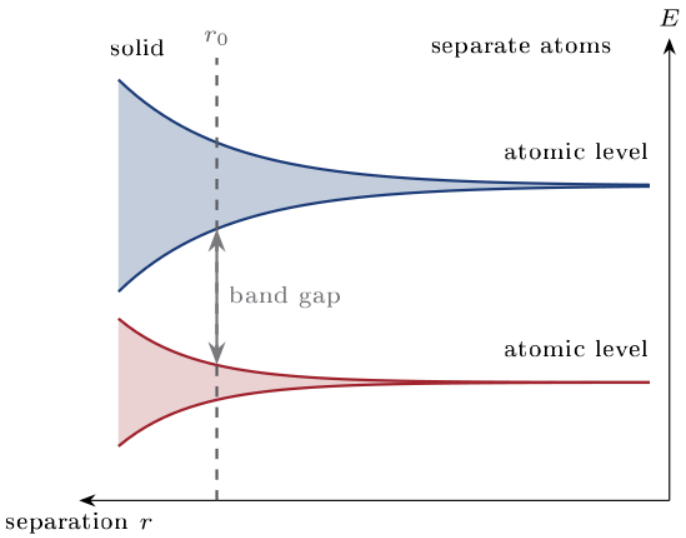


Figure 42.3. The formation of energy bands. Far apart (right), identical atoms share the same discrete levels. As they approach, each level splits into N closely spaced levels, one for each atom, forming a band. At the actual spacing r_0 of the solid, the bands are separated by a gap in which there are no electron states.

•Figure description

A diagram of electron energy against the separation between neighboring atoms, with separation decreasing to the left. At the right, where the atoms are far apart, there are two sharp horizontal lines: a lower atomic energy level and an upper one. Moving left, as the atoms approach, each line fans out into a shaded band whose width grows. The upper level, from the outer orbitals, starts spreading at a larger separation and becomes a wider band than the lower one. A vertical dashed line marks the actual spacing r sub 0 in the solid; at this spacing the two bands are separated by an empty energy range labeled band gap.

Each band holds $2N$ electrons for each atomic orbital it came from (the factor 2 is again spin), and the electrons fill the bands from the bottom up. Everything now depends on how the filling ends.

- **Conductors.** If the highest occupied band is only *partly* full, as in copper, there are empty states immediately above the filled ones. An applied field can nudge electrons into slightly higher-energy states moving in the direction of the force, and a current flows. The free electrons of Section 42.2 are the electrons of this partly filled band.
- **Insulators.** If the highest occupied band, the **valence band**, is completely full and the next band up, the **conduction band**, is separated from it by a wide gap, then no electron has an empty state within reach. A field cannot change any electron's motion, because every nearby state is taken, and no current flows. In diamond the gap is $5.5\ \text{eV}$, two hundred times k_BT at room temperature.
- **Semiconductors.** A semiconductor is an insulator with a small gap, about $1\ \text{eV}$ (Table 42.2). At absolute zero it is a perfect insulator, but at room temperature a few electrons are thermally excited across the gap into the conduction band, where they can carry current.

Figure 42.4 compares the three. Graphite and diamond differ in exactly this way. In diamond, each carbon atom uses all four valence electrons in covalent bonds, filling the valence band. In graphite, each carbon bonds to only three neighbors within a flat sheet, and the fourth electron of each atom spreads across the whole sheet in a band that is not full, so graphite conducts along its sheets almost like a metal.

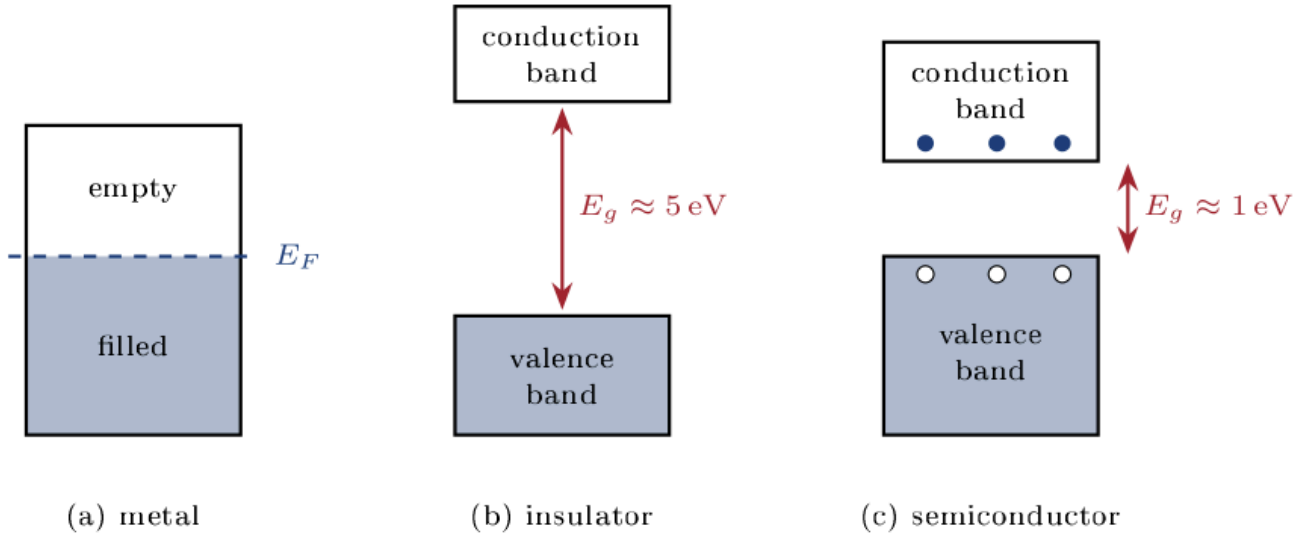


Figure 42.4. Band diagrams at room temperature. (a) In a metal the highest occupied band is only partly full, with empty states just above the Fermi energy. (b) In an insulator the valence band is full and a wide gap separates it from the empty conduction band. (c) In a semiconductor the gap is narrow enough that a few electrons (dots) are thermally excited into the conduction band, leaving holes (open circles) in the valence band.

•Figure description

Three side-by-side energy-band diagrams, energy increasing upward. Panel a, a metal: a single tall band, filled with electrons, shaded, only up to partway, with a dashed line labeled E sub F at the top of the filled region; empty states lie just above. Panel b, an insulator: a filled valence band at the bottom, a wide empty gap labeled E sub g about 5 eV, and an empty conduction band at the top. Panel c, a semiconductor: a filled valence band, a narrow gap labeled E sub g about 1 eV, and a conduction band containing a few electrons, shown as dots near its bottom, while the valence band has a few matching holes, shown as open circles near its top.

Table 42.2. Band gaps of some insulators and semiconductors at room temperature.

Material	E_g (eV)	Material	E_g (eV)
Quartz (SiO_2)	9	Gallium arsenide	1.42
Diamond	5.5	Silicon	1.12
Gallium nitride	3.4	Germanium	0.67
Silicon carbide	3.3	Indium antimonide	0.17

The number of electrons that heat can lift across a gap depends on a Boltzmann factor. For a pure material the density of conduction electrons is proportional to

$$e^{-E_g/2k_BT}, \tag{42.5}$$

where the factor of 2 arises because each excitation creates two carriers (as we will see shortly) and the energy cost is shared between them.

Example 42.5: Diamond versus silicon

Compare the factor $e^{-E_g/2k_BT}$ at $300\ \text{K}$ for (a) diamond and (b) silicon. (c) By what factor does silicon's value grow if it warms to $350\ \text{K}$?

Solution

At $300\ \text{K}$, $2k_BT = 2(8.617 \times 10^{-5})(300) = 0.0517\ \text{eV}$.

(a) Diamond: $e^{-5.5/0.0517} = e^{-106} \approx \boxed{6 \times 10^{-47}}$.

(b) Silicon: $e^{-1.12/0.0517} = e^{-21.7} \approx \boxed{4 \times 10^{-10}}$.

Silicon's factor is about 10^{37} times larger. A gap only five times smaller changes the result by thirty-seven powers of ten, because the gap sits in an exponent. This is why diamond is one of the best insulators known while silicon conducts measurably.

(c) At $350\ \mathrm{K}$, $2k_BT = 0.0603\ \mathrm{eV}$, so the factor becomes $e^{-1.12/0.0603} = e^{-18.6} = 8.6 \times 10^{-9}$, larger by

$$\frac{8.6 \times 10^{-9}}{3.9 \times 10^{-10}} \approx \boxed{22}. \tag{42.3.4}$$

A modest warming increases silicon's conductivity twentyfold. That is why the resistivity of a semiconductor *falls* steeply as it warms, the large negative temperature coefficient of silicon listed in Chapter 20, while a metal's resistivity rises.

Band gaps also decide how a solid looks. A photon can be absorbed by exciting an electron across the gap only if $hf \geq E_g$. Visible photons carry between 1.8 and $3.1\ \mathrm{eV}$, too little to cross diamond's $5.5\ \mathrm{eV}$ gap, so diamond lets visible light through and is transparent. Silicon's gap is smaller than any visible photon's energy, so silicon absorbs all visible light. In metals, with empty states immediately above the filled ones, electrons can absorb photons of any energy and re-radiate them, which gives metals their shine.

Checkpoint 42.4

Pure silicon is opaque to visible light but transparent to infrared light of wavelength $1.5\ \mu\mathrm{m}$, which is why it can be used for infrared lenses. Use $E_g = 1.12\ \mathrm{eV}$ and $hc = 1240\ \mathrm{eV \cdot nm}$ to explain both facts.

Answer

Answers to Checkpoints

Checkpoint 42.4. A photon is absorbed only if its energy is at least E_g, enough to lift an electron across the gap. The threshold wavelength for silicon is $\lambda = hc/E_g = 1240/1.12 \approx 1100\ \mathrm{nm}$. Every visible photon ($400$ –$700\ \mathrm{nm}$) has a shorter wavelength and more energy than this, so it is absorbed and silicon is opaque. A $1500\ \mathrm{nm}$ photon carries only $1240/1500 = 0.83\ \mathrm{eV}$, less than the gap, so it cannot be absorbed and passes through.

This page titled 42.3: Electrical Conduction in Metals and Insulators is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.4: Semiconductors

When an electron in a semiconductor is excited into the conduction band, it leaves behind an empty state in the valence band, called a **hole**. A hole behaves like a positive charge. When a field is applied, an electron in a neighboring bond can hop into the hole, which leaves a hole where that electron used to be; as the process repeats, the hole moves through the crystal in the direction a positive charge would move. A pure, or **intrinsic**, semiconductor has equal numbers of conduction electrons and holes, and both carry current. In pure silicon at room temperature there are only about 1×10^{16} of each per cubic meter, one for every five trillion atoms.

Doping. The properties of a semiconductor can be transformed by **doping**: adding a tiny, controlled amount of an impurity (Figure 42.5). Silicon has four valence electrons, each used in a bond to one of four neighbors. Replace one silicon atom with phosphorus, which has *five*. Four of phosphorus's electrons form the four bonds, and the fifth is left over, bound only weakly to the phosphorus ion. It takes only about $0.045\ \mathrm{eV}$ to free it into the conduction band, less than twice $k_B T$ at room temperature, so nearly every phosphorus atom donates an electron. Phosphorus is called a **donor**, and silicon doped this way is **n-type**, because its current is carried mostly by negative electrons. Replace a silicon atom with boron instead, which has only *three* valence electrons, and one bond is left an electron short: a hole. Boron is an **acceptor**, and silicon doped this way is **p-type**, with its current carried mostly by positive holes.

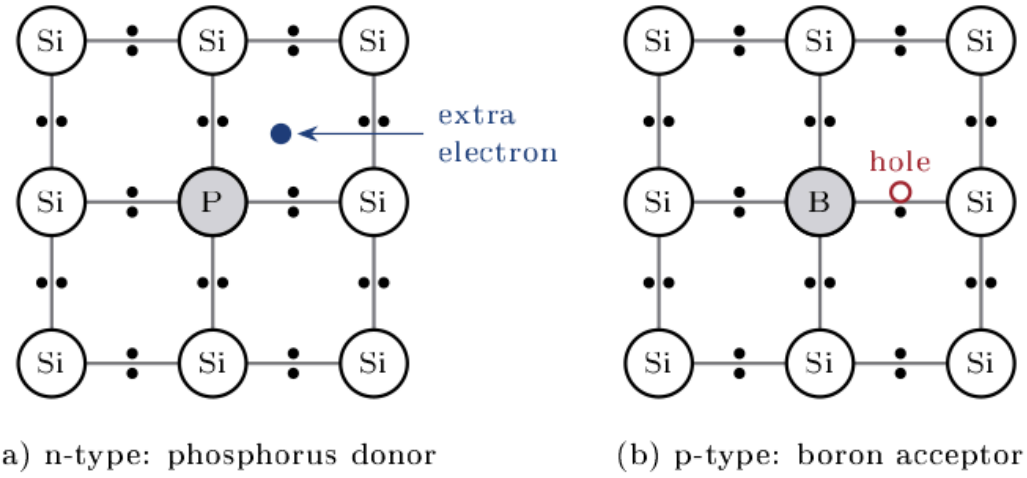


Figure 42.5. Doping silicon, shown flattened into two dimensions. Each bond holds a shared pair of electrons (dots). (a) A phosphorus atom has one more valence electron than it needs for its four bonds; the extra electron is easily freed into the conduction band. (b) A boron atom has one fewer, leaving a hole in one bond that a neighboring electron can hop into.

•Figure description

Two panels, each showing a flat, square grid of atoms as circles, joined to their four neighbors by bonds, with two small dots on each bond for the shared electron pair. Panel a, n-type: the central atom is phosphorus, labeled P, with five valence electrons; its four bonds are complete and a fifth, extra electron, drawn as a blue dot, sits loosely beside it, labeled extra electron. Panel b, p-type: the central atom is boron, labeled B, with three valence electrons; one of its four bonds has only one dot and an open circle in place of the second, labeled hole.

Example 42.6: A pinch of phosphorus

Silicon has $5.0 \times 10^{28}\ \mathrm{atoms/m^3}$. A sample is doped with one phosphorus atom for every 10^7 silicon atoms, and each phosphorus atom donates one conduction electron. Compare the density of conduction electrons with that of pure silicon, $1.0 \times 10^{16}\ \mathrm{m^{-3}}$.

Solution

The donor density is

$$n_d = \frac{5.0 \times 10^{28}}{10^7} = 5.0 \times 10^{21}\,\mathrm{m}^{-3}, \tag{42.4.1}$$

so the conduction-electron density rises by a factor of

$$\frac{5.0 \times 10^{21}}{1.0 \times 10^{16}} = \boxed{5 \times 10^5}. \tag{42.4.2}$$

A doping level of one part in ten million, far below the impurity level of any “pure” chemical sold in a store, increases the conductivity half a million times. That is why semiconductor manufacturing demands silicon purified to better than one part in a billion before the dopants are added deliberately, and why the dopants themselves are placed with such care.

Checkpoint 42.5

Germanium, like silicon, has four valence electrons. Is germanium doped with arsenic (five valence electrons) n-type or p-type? What about germanium doped with gallium (three)?

Answer

Diodes

The real power of doping appears when p-type and n-type material meet in a single crystal, forming a **p–n junction** (Figure 42.6a). Electrons from the n side diffuse across the boundary and fill holes on the p side. This leaves a thin layer on the n side with positive donor ions and no free electrons, and a layer on the p side with negative acceptor ions and no holes. This layer, depleted of mobile carriers, is the **depletion zone**. Its fixed charges create an electric field pointing from the n side to the p side, and a potential difference of about $0.7\ \mathrm{V}$ in silicon, that stops further diffusion.

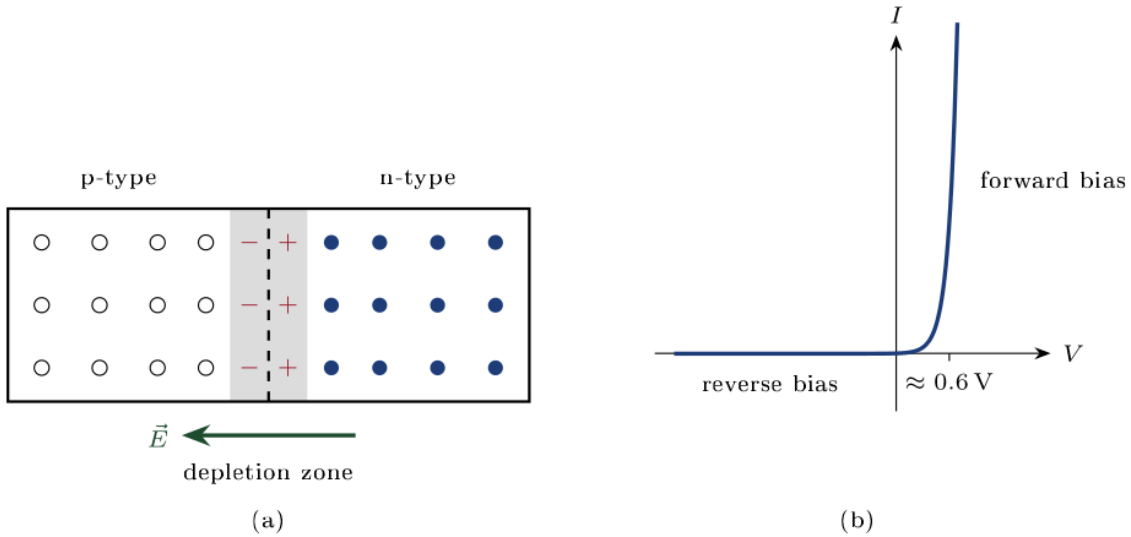


Figure 42.6. (a) A p–n junction. Electrons diffusing from the n side fill holes on the p side, leaving a depletion zone of fixed ions whose electric field opposes further diffusion. (b) The current through a junction diode. Under reverse bias almost no current flows; under forward bias the current grows exponentially and becomes large beyond about $0.6\ \mathrm{V}$ in silicon.

•Figure description

Two panels. Panel a, a p-n junction: a rectangular crystal, p-type on the left and n-type on the right. Open circles for holes fill the p side and dots for electrons fill the n side, except in a narrow shaded strip at the boundary, the depletion zone. Inside the strip, minus signs mark fixed negative acceptor ions on the p side and plus signs mark fixed positive donor ions on the n side, and an arrow labeled E points from the n side to the p side. Panel b, the current-voltage curve of a diode: for negative, reverse-bias voltages the current is essentially zero; for positive, forward-bias voltages it stays near zero until about 0.6 volts and then rises extremely steeply.

Now connect the junction to a battery. If the p side is made positive, called **forward bias**, the applied voltage opposes the built-in field and shrinks the depletion zone. Electrons pour across from the n side and holes from the p side, and a large current flows. If the p side is made negative, called **reverse bias**, the applied voltage adds to the built-in field, the depletion zone widens, and almost no current flows. The junction is a **diode**: a one-way valve for current. The current through an ideal diode at voltage V is

Key Equation: The diode equation

$$I = I_0\left(e^{eV/k_BT} - 1\right), \tag{42.6}$$

where I_0 is a small saturation current, typically 10^{-12} amperes or less for a silicon diode. For reverse bias ($V < 0$) the exponential quickly vanishes and $I \to -I_0$, a negligible trickle. For forward bias the current grows by a factor of e for every $k_BT/e = 25.9\ \text{mV}$ at room temperature (Figure 42.6b). Diodes convert the alternating current of Chapter 26 to direct current in every phone charger.

Example 42.7: Turning on a diode

A silicon diode at $300\ \text{K}$ has $I_0 = 1.0 \times 10^{-12}\ \text{A}$. Find the current at (a) $V = +0.60\ \text{V}$, (b) $V = +0.66\ \text{V}$, and (c) $V = -5.0\ \text{V}$.

Solution

At $300\ \text{K}$, $k_BT/e = 0.02585\ \text{V}$.

(a) $eV/k_BT = 0.60/0.02585 = 23.2$, so

$$I = (1.0 \times 10^{-12})(e^{23.2} - 1) = (1.0 \times 10^{-12})(1.2 \times 10^{10}) = \boxed{12\,\text{mA}}. \tag{42.4.3}$$

(b) $eV/k_BT = 0.66/0.02585 = 25.5$, so

$$I = (1.0 \times 10^{-12})(1.2 \times 10^{11}) = \boxed{0.12\,\text{A}}. \tag{42.4.4}$$

An increase of only $60\ \text{mV}$ multiplies the current by ten. Over a wide range of currents, a forward-biased silicon diode therefore holds a nearly constant voltage of about 0.6 to $0.7\ \text{V}$.

(c) $e^{-5.0/0.02585}$ is utterly negligible, so $I = \boxed{-1.0 \times 10^{-12}\,\text{A}}$: a picoampere, eleven orders of magnitude smaller than in (a).

Light-emitting diodes and solar cells. In a forward-biased junction, electrons crossing into the p side drop into holes, falling across the band gap. In silicon the energy goes mostly into lattice vibrations. But in some semiconductors, such as gallium arsenide and gallium nitride, each such recombination usually emits a photon with energy close to E_g. A junction made of such a material is a **light-emitting diode**, or LED, and its color is set by its band gap:

$$\lambda \approx \frac{hc}{E_g}. \tag{42.7}$$

Gallium arsenide ($E_g = 1.42\ \text{eV}$) emits infrared at $870\ \text{nm}$, used in television remote controls. Blending in phosphorus raises the gap to give red LEDs, and the gallium nitride family yields blue and violet. The blue LED, whose invention earned the 2014 Nobel Prize in Physics, made white LED lighting possible; a white LED is a blue LED coated with a phosphor that converts part of its light to yellow. Put a pair of mirrors on the ends of a suitable junction and the stimulated emission of Section 41.7 turns the LED into a **diode laser**, the laser in every barcode scanner and fiber-optic link.

Run the process backward and the diode becomes a **solar cell**. A photon with $hf \geq E_g$ absorbed near the junction creates an electron–hole pair, and the built-in field sweeps the electron to the n side and the hole to the p side. Charge builds up on each side and drives a current through an external circuit. Any photon energy beyond E_g is lost as heat, and photons with less than E_g pass through unused, which is why there is an optimum band gap for converting sunlight, close to silicon's.

The Transistor

A diode lets current flow one way. A **transistor** lets a small signal *control* a larger current, and that makes it both an amplifier and a switch. The first transistor was built by John Bardeen, Walter Brattain, and William Shockley at Bell Laboratories in 1947. The

kind used in nearly all modern electronics is the **metal–oxide–semiconductor field-effect transistor**, or MOSFET (Figure 42.7).

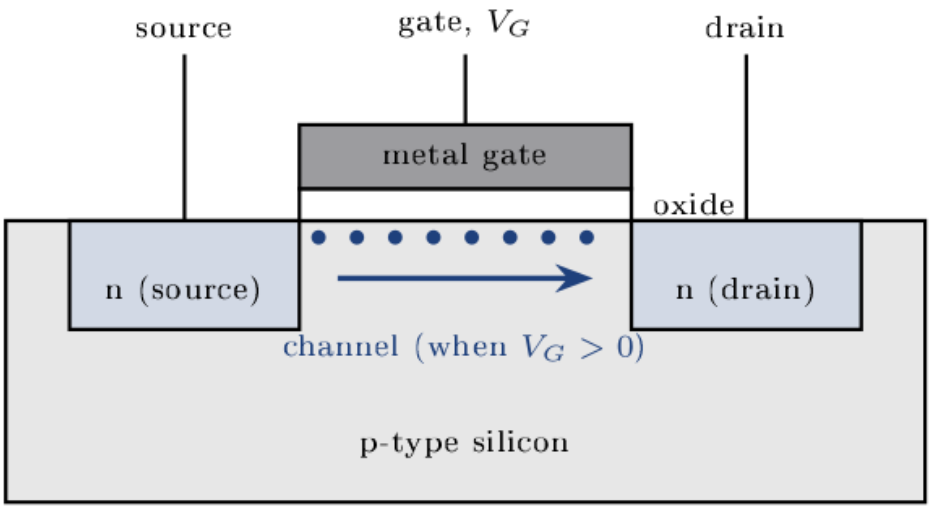


Figure 42.7. An n-channel MOSFET. With no gate voltage, the p-type region between source and drain blocks current. A positive gate voltage attracts electrons to the surface beneath the thin oxide, forming a conducting n-type channel between source and drain. The gate draws essentially no current, since the oxide insulates it.

•Figure description

A cross-section of a MOSFET. A block of p-type silicon forms the base. Set into its top surface are two n-type regions, the source on the left and the drain on the right, each with a wire leading up to its terminal. Between them, on top of the silicon, lies a thin insulating layer of silicon dioxide, and on top of that a metal gate electrode with its own terminal, labeled V sub G. Just beneath the oxide, connecting source and drain, a thin layer of electrons drawn as dots is labeled conducting channel, formed when V sub G is positive. An arrow shows electrons flowing from the source through the channel to the drain.

Two n-type regions, the **source** and the **drain**, are set into a block of p-type silicon. Between them, separated from the silicon by a layer of insulating oxide only a few nanometers thick, is a metal electrode called the **gate**. With no voltage on the gate, current cannot pass from source to drain: the path crosses p-type silicon, and whichever way the voltage is applied, one of the two p–n junctions along the path is reverse biased. Now make the gate positive. The gate and the silicon beneath it form a capacitor (Chapter 19), and the positive gate attracts electrons to the surface just under the oxide. Once the gate voltage is large enough, these electrons outnumber the holes in a thin layer, which becomes effectively n-type: a conducting **channel** connects source and drain, and current flows. The gate voltage controls the current, but because the oxide insulates it, the gate itself draws almost no current, so a tiny amount of power controls a much larger one. Used with a range of gate voltages, the MOSFET is an **amplifier**; switched between “off” and “on,” it is a **switch**, and a switch is all that is needed to represent the 0s and 1s of binary logic.

Checkpoint 42.6

A p-channel MOSFET is built the other way around: p-type source and drain set into n-type silicon. Should its gate be made positive or negative to turn it on, and what kind of carrier then fills the channel?

Answer

The Integrated Circuit

Early transistors were wired together one at a time, like the resistors and capacitors of a circuit board. In 1958 and 1959, Jack Kilby at Texas Instruments and Robert Noyce at Fairchild Semiconductor independently showed that transistors, resistors, capacitors, and the connections between them could all be built at once into a single chip of silicon: the **integrated circuit**. The patterns are printed by **photolithography**: a light-sensitive coating on the wafer is exposed to ultraviolet light through a mask, the exposed regions are etched away, dopants are implanted, and the cycle is repeated dozens of times to build up the layers of a circuit.

Because every transistor on a wafer is made in the same steps, shrinking the patterns puts more transistors on a chip at almost no

extra cost. In 1965 Gordon Moore observed that the number of transistors on a chip was doubling at a steady rate; the doubling time settled at about two years and held for half a century, a trend known as **Moore's law**.

Example 42.8: Fifty years of Moore's law

The first commercial microprocessor, the Intel 4004 of 1971, held about 2300 transistors. If the number doubled every 2.0 years, how many transistors would a chip hold in 2023?

Solution

From 1971 to 2023 is 52 years, or 26 doublings, so

$$N = 2300 \times 2^{26} = 2300 \times (6.7 \times 10^{7}) = \boxed{1.5 \times 10^{11}}. \tag{42.4.5}$$

The largest processors of 2023 did indeed contain somewhat more than 10^{11} transistors. Fifty years of exponential growth multiplied the count by nearly a hundred million.

The growth cannot go on forever. The smallest features of today's chips are only a few tens of atoms across, and the gate oxide is only a few atoms thick. At that scale the quantum effects that made the transistor possible begin to work against it: electrons tunnel straight through the thin oxide (Chapter 40), leaking current and wasting power. Much of modern chip design is a struggle against tunneling, waged with new materials and three-dimensional transistor shapes. The phones, computers, and networks built from these chips are the most tangible products of the quantum mechanics of the last four chapters.

Answers to Checkpoints

Checkpoint 42.5. Arsenic has one more valence electron than germanium needs for its four bonds, so it donates an extra electron: arsenic-doped germanium is *n-type*. Gallium has one fewer, leaving a hole: gallium-doped germanium is *p-type*.

Checkpoint 42.6. Negative. Everything is reversed: a negative gate repels electrons from the n-type surface and attracts holes, and once enough holes gather under the oxide, they form a p-type channel joining the p-type source and drain. The current in the channel is carried by holes. (Modern chips pair n-channel and p-channel transistors in "complementary" (CMOS) circuits, so that in either logic state one of each pair is off and almost no current flows.)

This page titled 42.4: Semiconductors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.S: Bonding and Electrical Conduction in Solids (Summary)

Summary

- A **bond** forms when two atoms have lower energy together than apart. The potential energy $U(r)$ has a minimum at the equilibrium separation r_0; the well depth is the dissociation energy. Short-range repulsion comes mostly from the Pauli exclusion principle.
- **Ionic** bonds join ions formed by electron transfer ($ke^2 = 1.44\ \text{eV}\cdot\text{nm}$); **covalent** bonds share a spin-paired electron pair; **metallic** bonds immerse ions in a sea of electrons; **hydrogen** and **van der Waals** bonds are much weaker.
- Molecular vibrations are quantized: $E_v = (v + \frac{1}{2})hf$ with $f = (1/2\pi)\sqrt{k/\mu}$. When $hf \gg k_BT$ the vibration is frozen out.
- In the **free-electron model**, a metal's valence electrons fill the states of a 3D box, two per state, up to the **Fermi energy** $E_F = (h^2/2m)(3n/8\pi)^{2/3}$, about $7\ \text{eV}$ in copper. The **Fermi–Dirac distribution** $f(E) = 1/(e^{(E-E_F)/k_BT} + 1)$ shows that only electrons within a few k_BT of E_F are affected by temperature.
- A metal's resistivity is $\rho = m/(ne^2\tau)$; electron waves pass through a perfect lattice and scatter only from imperfections and vibrations.
- The levels of N atoms in a solid broaden into **energy bands** separated by **gaps**. A partly filled band makes a **conductor**; a full valence band with a wide gap makes an **insulator**; a gap of about $1\ \text{eV}$ makes a **semiconductor**, whose carrier density varies as $e^{-E_g/2k_BT}$.
- **Doping** creates n-type (donor, extra electrons) or p-type (acceptor, holes) semiconductors. A **p–n junction** is a diode, $I = I_0(e^{eV/k_BT} - 1)$; LEDs emit at $\lambda \approx hc/E_g$, and solar cells run the process in reverse.
- In a **MOSFET**, a gate voltage creates a conducting channel between source and drain, making a switch or amplifier. **Integrated circuits** hold more than 10^{11} transistors, the product of decades of exponential shrinking.

Key Equations

Coulomb energy of an ion pair

$$U = -\frac{ke^2}{r}, \qquad ke^2 = 1.44\,\text{eV}\cdot\text{nm} \tag{42.S.1}$$

Vibrational frequency and levels

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}}, \quad \mu = \frac{m_1 m_2}{m_1 + m_2}, \qquad E_v = \left(v + \tfrac{1}{2}\right) hf \tag{42.S.2}$$

Fermi energy of a metal

$$E_F = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3}, \qquad v_F = \sqrt{\frac{2E_F}{m}} \tag{42.S.3}$$

Fermi–Dirac occupation probability

$$f(E) = \frac{1}{e^{(E-E_F)/k_BT} + 1} \tag{42.S.4}$$

Resistivity of a metal

$$\rho = \frac{m}{ne^2\tau} \tag{42.S.5}$$

Carriers thermally excited across a gap

$$n \propto e^{-E_g/2k_BT} \tag{42.S.6}$$

The diode equation

$$I = I_0\left(e^{eV/k_BT} - 1\right) \tag{42.S.7}$$

LED wavelength

$$\lambda \approx \frac{hc}{E_g} \tag{42.S.8}$$

This page titled 42.S: Bonding and Electrical Conduction in Solids (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

42.E: Bonding and Electrical Conduction in Solids (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use $h = 6.626 \times 10^{-34}\ \text{J}\cdot\text{s}$, $m_e = 9.109 \times 10^{-31}\ \text{kg}$, $e = 1.602 \times 10^{-19}\ \text{C}$, $k_B = 8.617 \times 10^{-5}\ \text{eV/K} = 1.381 \times 10^{-23}\ \text{J/K}$, $hc = 1240\ \text{eV}\cdot\text{nm}$, $ke^2 = 1.44\ \text{eV}\cdot\text{nm}$, $1\ \text{u} = 1.661 \times 10^{-27}\ \text{kg}$, and Tables 42.1 and 42.2.

Section 42.1 — Molecular Bonds

42.1 ★ *[Conceptual.]* Name the main type of bond holding together each of the following: (a) a crystal of table salt; (b) a hydrogen molecule; (c) solid argon; (d) one water molecule to its neighbor in ice; (e) a copper wire.

42.2 ★ *[Conceptual.]* Two hydrogen atoms bond to form H_2, but two helium atoms do not form a stable He_2 molecule. Use the Pauli exclusion principle to explain the difference.

42.3 ★★ Potassium chloride, KCl, is an ionic molecule with a bond length of $0.267\ \text{nm}$. The ionization energy of potassium is $4.34\ \text{eV}$, and the electron affinity of chlorine is $3.61\ \text{eV}$. (a) Find the net energy needed to transfer an electron from K to Cl when the atoms are far apart. (b) Estimate the energy released in forming the molecule, ignoring the short-range repulsion. (c) At what separation does the electron transfer first become favorable?

42.4 ★★ The van der Waals bond between two argon atoms has an energy of about $0.010\ \text{eV}$. At what temperature does k_BT equal this energy? Compare with argon's boiling point, $87\ \text{K}$, and comment.

42.5 ★★ *[Real-world application: infrared spectroscopy.]* The HCl molecule absorbs infrared light of wavelength $3.46\ \mu\text{m}$ when it jumps from the $v = 0$ to the $v = 1$ vibrational level. The atomic masses are $1.008\ \text{u}$ for H and $35.45\ \text{u}$ for Cl. (a) Find the vibrational quantum hf in eV, and the frequency f. (b) Find the reduced mass and the effective spring constant of the bond. (c) What fraction of HCl molecules are in the $v = 1$ level at $300\ \text{K}$?

42.6 ★★★ The potential well of the hydrogen molecule H_2 is $4.75\ \text{eV}$ deep, and its vibrational quantum is $hf = 0.546\ \text{eV}$. (a) Even at absolute zero the molecule has its zero-point energy. How much energy is needed to break apart a ground-state H_2 molecule? (b) The molecule D_2 is made of deuterium atoms, which have twice the mass of hydrogen atoms but the same electrons, so the same potential well and the same spring constant. Find its vibrational quantum and its dissociation energy. (c) Which molecule is harder to break apart, and why?

Section 42.2 — Free-Electron Theory of Metals

42.7 ★ Gold has one free electron per atom and $n = 5.90 \times 10^{28}\ \text{m}^{-3}$. Find (a) its Fermi energy, (b) its Fermi speed, and (c) its Fermi temperature.

42.8 ★★ Aluminum has three free electrons per atom, a density of $2700\ \text{kg/m}^3$, and a molar mass of $26.98\ \text{g/mol}$. (a) Find the density of free electrons. (b) Find the Fermi energy, and compare it with copper's $7.0\ \text{eV}$.

42.9 ★★ For a metal at $300\ \text{K}$, find the occupation probability $f(E)$ at (a) $E = E_F + 0.050\ \text{eV}$, (b) $E = E_F - 0.050\ \text{eV}$, and (c) $E = E_F + 0.25\ \text{eV}$. (d) How far above E_F does the occupation probability fall to 0.010?

42.10 ★★ Show that the Fermi–Dirac distribution is symmetric about the Fermi energy in the sense that $f(E_F + \Delta) + f(E_F - \Delta) = 1$ for any Δ. Interpret the result in terms of electrons above E_F and empty states below it.

42.11 ★★ At absolute zero the average kinetic energy of the free electrons in a metal is $\frac{3}{5}E_F$. (a) Find this average energy for copper. (b) Compare it with the average translational kinetic energy of a gas molecule at $300\ \text{K}$. (c) At what temperature would a classical gas have an average kinetic energy equal to the answer to (a)?

42.12 ★★ *[Conceptual.]* Classical physics predicts that the free electrons in a metal should add $\frac{3}{2}k_B$ apiece to its heat capacity, but they add almost nothing at room temperature. Explain why, and estimate the fraction of copper's free electrons that can absorb thermal energy at $300\ \text{K}$.

Section 42.3 — Electrical Conduction in Metals and Insulators

42.13 ★ *[Conceptual.]* A completely full band holds a great many electrons, yet it carries no current when a field is applied. Explain why.

42.14 ★★ Silver has a resistivity of $1.59 \times 10^{-8}\ \Omega \cdot \mathrm{m}$, a free-electron density of $5.86 \times 10^{28}\ \mathrm{m}^{-3}$, and a Fermi speed of $1.39 \times 10^{6}\ \mathrm{m/s}$. Find (a) the mean time between collisions and (b) the mean free path of its conduction electrons.

42.15 ★★ Visible light spans photon energies from about $1.8\ \mathrm{eV}$ (red) to $3.1\ \mathrm{eV}$ (violet). Using Table 42.2, decide which of diamond, gallium nitride, silicon, and germanium are transparent to visible light. For gallium arsenide, find the longest wavelength it can absorb.

42.16 ★★ Using the factor $e^{-E_g/2k_BT}$ at $300\ \mathrm{K}$, estimate the ratio of the number of conduction electrons in pure germanium to that in pure silicon. Compare with the ratio of their resistivities, $2.3 \times 10^{3}\ \Omega \cdot \mathrm{m}$ for silicon and $0.46\ \Omega \cdot \mathrm{m}$ for germanium, and comment.

42.17 ★★ *[Conceptual.]* The resistance of a copper wire rises as it warms, while the resistance of a piece of pure silicon falls. Explain each behavior in terms of $\rho = m/(ne^2\tau)$: which factor changes most with temperature in each material, and why?

Section 42.4 — Semiconductors, Diodes, Transistors, and Integrated Circuits

42.18 ★ *[Conceptual.]* Silicon is doped with each of the following. State whether the result is n-type or p-type: (a) arsenic (5 valence electrons); (b) gallium (3); (c) boron (3); (d) antimony (5).

42.19 ★★ Silicon ($5.0 \times 10^{28}\ \mathrm{atoms/m^3}$) is doped with one boron atom for every 10^{6} silicon atoms. (a) Find the density of holes contributed by the boron. (b) By what factor does this exceed the hole density in pure silicon, $1.0 \times 10^{16}\ \mathrm{m}^{-3}$?

42.20 ★★ A silicon diode has $I_0 = 2.0 \times 10^{-12}\ \mathrm{A}$. (a) What forward voltage gives a current of $10\ \mathrm{mA}$ at $300\ \mathrm{K}$? (b) Keeping that voltage and ignoring any change in I_0, find the current at $320\ \mathrm{K}$. (In real diodes I_0 grows rapidly with temperature, which more than reverses this effect.)

42.21 ★★ *[Real-world application: LEDs and solar cells.]* (a) Estimate the band gaps of the semiconductors in a red LED emitting at $630\ \mathrm{nm}$ and a blue LED emitting at $450\ \mathrm{nm}$. (b) What is the longest wavelength that a silicon solar cell can convert? (c) A $500\ \mathrm{nm}$ photon is absorbed in a silicon solar cell. What fraction of its energy is lost as heat, at the least?

42.22 ★★ *[Real-world application: computing.]* (a) Starting from 2300 transistors in 1971 and doubling every 2.0 years, how many transistors would a chip hold in 2021? (b) If the trend continued to 2041, how many would a chip hold? (c) If those 2041 transistors had to fit on a chip $1.0\ \mathrm{cm}^2$ in area, find the side of the square area available to each, and compare it with the spacing of silicon atoms, $0.235\ \mathrm{nm}$. What does this suggest?

Challenge Problems

42.23 ★★★ *[A hydrogen atom inside a crystal.]* The extra electron of a phosphorus donor in silicon is bound to the $+e$ charge of the phosphorus ion, much like the electron of a hydrogen atom. Two things change. The surrounding silicon reduces the electric force by its dielectric constant, $\kappa = 11.7$ (Chapter 19), and the electron moves through the crystal as if its mass were only $m^* = 0.26\ m_e$.

a. In the Bohr results of Chapter 41, $E_1 = -me^4/(8\epsilon_0^2 h^2)$ and $a_0 = \epsilon_0 h^2/(\pi m e^2)$. Show that the replacements $\epsilon_0 \to \kappa\epsilon_0$ and $m \to m^*$ give a binding energy of $(13.6\ \mathrm{eV})(m^*/m_e)/\kappa^2$ and an orbit radius of $a_0\kappa(m_e/m^*)$.
b. Evaluate the binding energy and the radius. Compare the binding energy with k_BT at $300\ \mathrm{K}$, and the radius with the spacing of silicon atoms, $0.235\ \mathrm{nm}$.
c. The measured binding energy is $0.045\ \mathrm{eV}$. Using your results, explain why nearly every donor is ionized at room temperature, and why treating the silicon as a smooth dielectric is a reasonable approximation.

42.24 ★★★ *[Real-world application: materials — the chapter callback.]* Reread the chapter opener. The density of

conduction electrons in a pure semiconductor or insulator is roughly $n \approx N_{\text{eff}}\ e^{-E_g/2k_BT}$, with $N_{\text{eff}} \approx 2.5 \times 10^{25}\ \text{m}^{-3}$ at room temperature.

a. Estimate n for pure silicon at $300\ \text{K}$ and compare it with the measured $1.0 \times 10^{16}\ \text{m}^{-3}$.
b. Estimate n for diamond at $300\ \text{K}$. What volume of diamond would contain, on average, a single conduction electron? Express the answer as the side of a cube, and compare it with the size of Earth (diameter $1.27 \times 10^{7}\ \text{m}$).
c. Copper has 8.49×10^{28} free electrons per cubic meter. Find the ratio of copper's carrier density to silicon's, and compare it with the ratio of their resistivities, $2.3 \times 10^{3}\ \Omega \cdot \text{m}$ and $1.68 \times 10^{-8}\ \Omega \cdot \text{m}$. Using $\rho = m/(ne^2\tau)$, what must be true of the collision time τ in silicon compared with copper?
d. Diamond and graphite are both pure carbon. Explain, in terms of bonding and bands, why graphite conducts and diamond does not.

This page titled 42.E: Bonding and Electrical Conduction in Solids (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43: NUCLEAR PHYSICS

Every second, the Sun turns about four million tonnes of its own mass into light. A single kilogram of uranium, fully fissioned, releases as much energy as burning roughly three thousand tonnes of coal. A gram of matter, if it could be converted entirely to energy, would power a small town for weeks. These staggering numbers all come from the same place: the tiny, dense core of the atom, and Einstein's relation $E = mc^2$, which says that mass *is* a form of energy.

Remarkably, both the Sun's fusion and a reactor's fission are explained by a single graph: the curve of binding energy per nucleon. It tells us that nuclei near iron are the most tightly bound of all, and that energy is released whenever nuclei move *toward* iron: light nuclei by fusing together, heavy nuclei by splitting apart. The same curve explains why stars shine, why some elements are radioactive, where the chemical elements came from, and why a nuclear weapon can level a city. In this chapter we build up the physics of the nucleus from its size and composition, through binding energy and radioactive decay, to the great energy-releasing processes of fission and fusion. By the end you will be able to predict how a nucleus decays, calculate the energy a reaction releases, and understand why a kilogram of nuclear fuel so vastly outperforms a kilogram of anything you can burn.

Learning Objectives

- Describe a nucleus by its atomic number, neutron number, and mass number, and estimate its size.
- Compute the binding energy and binding energy per nucleon of a nucleus, and read the curve of binding energy.
- Use the exponential decay law, the decay constant, and half-life to find activities and ages, including radiocarbon dating.
- Balance and find the energy released (Q value) of nuclear reactions, including alpha and beta decay.
- Explain nuclear fission and the chain reaction that powers reactors, and compute the energy released.
- Explain nuclear fusion, the Coulomb barrier, and the proton–proton chain that powers the Sun.

We will return to the binding-energy curve again and again, and in the chapter-callback problem near the end you will compare the energy stored in a kilogram of chemical fuel, fission fuel, and fusion fuel.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

43.1: Properties of Nuclei
43.2: Nuclear Binding Energy
43.3: Radioactivity
43.4: Nuclear Reactions
43.S: Nuclear Physics (Summary)
43.E: Nuclear Physics (Problems)

This page titled 43: Nuclear Physics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.1: Properties of Nuclei

A nucleus is built from two kinds of **nucleons**: positively charged *protons* and uncharged *neutrons*. Three integers describe it. The **atomic number** Z is the number of protons (and fixes which element it is); the **neutron number** N is the number of neutrons; and the **mass number** A is the total number of nucleons:

Key Equation: Mass number

$$A = Z + N. \tag{43.1}$$

A specific nuclear species, or **nuclide**, is written ${}^{A}_{Z}\mathrm{X}$, where X is the chemical symbol. For example, ${}^{14}_{6}\mathrm{C}$ has $Z = 6$ protons and $N = 14 - 6 = 8$ neutrons. Because the element is fixed by Z alone, the symbol and the subscript are redundant; ${}^{14}_{6}\mathrm{C}$ and "carbon-14" carry the same information.

Atoms of one element that differ only in neutron number are called **isotopes**. Carbon, for instance, occurs as the stable ${}^{12}\mathrm{C}$ and ${}^{13}\mathrm{C}$ and the radioactive ${}^{14}\mathrm{C}$: all six-proton carbon, but with six, seven, and eight neutrons respectively.

Nuclei are astonishingly small. Their radii are measured in **femtometers** (also called fermis), $1\ \mathrm{fm} = 10^{-15}\ \mathrm{m}$, and to good accuracy the radius depends only on the number of nucleons:

Key Equation: Nuclear radius

$$r = aA^{1/3}, \qquad a = 1.2\,\mathrm{fm}. \tag{43.2}$$

The cube-root dependence has a profound consequence: since volume scales as $r^3 \propto A$, every nucleus packs the same number of nucleons into a given volume. *All nuclei have essentially the same density*: about $2.3 \times 10^{17}\ \mathrm{kg/m^3}$, the density of the nucleons themselves.

Because nuclear masses are so small, we measure them in **atomic mass units**, defined so that one neutral ${}^{12}\mathrm{C}$ atom has a mass of exactly $12\ \mathrm{u}$:

$$1\,\mathrm{u} = 1.660539 \times 10^{-27}\,\mathrm{kg}. \tag{43.3}$$

Through $E = mc^2$, mass and energy are interchangeable, and the most useful conversion in all of nuclear physics is

$$c^2 = 931.494\,\mathrm{MeV/u}, \tag{43.4}$$

meaning that $1\ \mathrm{u}$ of mass is equivalent to $931.494\ \mathrm{MeV}$ of energy.

Example 43.1: The size and density of a nucleus

Estimate the radius of a ${}^{238}\mathrm{U}$ nucleus, and compare it to that of a ${}^{4}\mathrm{He}$ nucleus.

Solution

Using Equation (43.2),

$$r_{\mathrm{U}} = (1.2\,\mathrm{fm})(238)^{1/3} = (1.2)(6.20)\,\mathrm{fm} = \boxed{7.4\,\mathrm{fm}}, \tag{43.1.1}$$

while

$$r_{\mathrm{He}} = (1.2\,\mathrm{fm})(4)^{1/3} = (1.2)(1.59)\,\mathrm{fm} = \boxed{1.9\,\mathrm{fm}}. \tag{43.1.2}$$

Uranium has $238/4 \approx 60$ times as many nucleons as helium, yet its radius is only $(60)^{1/3} \approx 3.9$ times larger, exactly because $r \propto A^{1/3}$. The two nuclei have the same density.

What holds all those protons together against their mutual electrical repulsion? Not gravity, which is far too weak, and certainly not the electric force, which pushes the protons apart. The answer is a third fundamental interaction, the **strong nuclear force**. At the scale of a femtometer it is about 137 times stronger than the electromagnetic force, and overwhelmingly stronger than gravity. Crucially, it is *short-ranged*: each nucleon attracts only its nearest neighbors, over distances of about $1\ \text{fm}$, and beyond that the force essentially vanishes. We will look more closely at the strong interaction in Chapter 44.

Checkpoint 43.1

Two nuclei, one with $A = 16$ and one with $A = 128$, are compared. The second has eight times as many nucleons. By what factor is its radius larger, and how do the two nuclear densities compare? Explain using $r \propto A^{1/3}$.

Answer

Answers to Checkpoints

Checkpoint 43.1. Since $r \propto A^{1/3}$, eight times the nucleons gives a radius $8^{1/3} = 2$ times larger. The volume goes as $r^3 \propto A$, so it grows eightfold too: eight times the mass in eight times the volume. The densities are the same. Nucleons pack together at essentially a fixed density, whatever the size of the nucleus.

This page titled 43.1: Properties of Nuclei is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.2: Nuclear Binding Energy

Here is a startling fact: the mass of any nucleus is always *less* than the combined mass of its separate protons and neutrons. If you could weigh a helium nucleus and then weigh its two protons and two neutrons separately, the loose parts would weigh more. Writing M for the mass of the assembled nucleus and $\sum m$ for the sum of the masses of its individual nucleons,

$$M < \sum m. \tag{43.5}$$

The missing mass has not vanished; by $E = mc^2$ it has been released as energy when the nucleus formed. That same energy must be supplied to pull the nucleus apart again, so it is called the **binding energy** E_b:

Key Equation: Nuclear binding energy

$$E_b = \left(\sum m - M\right) c^2 = \left(\sum m\right) c^2 - Mc^2. \tag{43.6}$$

The quantity $\sum m - M$ is the **mass defect**. In practice we compute $\sum m$ from tabulated atomic masses (the electron masses cancel for a neutral atom), multiply the defect by $c^2 = 931.494\ \text{MeV/u}$, and obtain E_b in MeV.

A more revealing measure is the binding energy divided by the number of nucleons, the **binding energy per nucleon**:

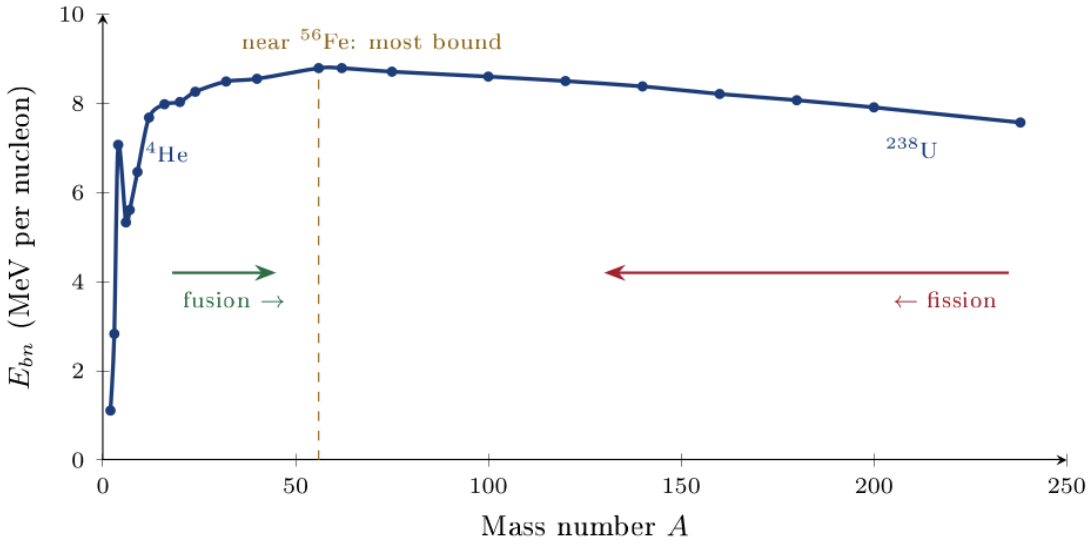


Figure 43.1. The curve of binding energy per nucleon, E_{bn}, versus mass number A. The peak near ^{56}Fe (about $8.8\ \text{MeV}$) marks the most tightly bound nuclei. Energy is released whenever nucleons end up more tightly bound, by *fusing* light nuclei (moving right, up the steep left side) or by *splitting* heavy ones (moving left, up the gentle right side). Both processes climb toward the iron peak.

•Figure description

A graph of binding energy per nucleon in MeV against mass number A, from A equals 2 to about 240. The curve climbs very steeply from about 1.1 MeV at A equals 2, through a spike at helium-4 near 7.1, and reaches a broad flat maximum of about 8.8 MeV around A equals 56, marked by a dashed line labeled near iron-56, most bound. Beyond the peak it declines slowly and steadily, reaching about 7.6 MeV at uranium-238. A green arrow on the steep left-hand side points right, labeled fusion; a red arrow on the gentle right-hand side points left, labeled fission. Both arrows climb toward the iron peak, which is why both processes release energy.

Key Equation: Binding energy per nucleon

$$E_{bn} = \frac{E_b}{A}. \tag{43.7}$$

The larger E_{bn} is, the more tightly each nucleon is held, and the more stable the nucleus. Plotting E_{bn} against mass number gives one of the most important graphs in physics (Figure 43.1). It rises steeply for the lightest nuclei, reaches a broad maximum of about $8.8\ \text{MeV}$ near ^{56}Fe and ^{62}Ni, then declines slowly toward the heaviest nuclei. The nuclei near iron are the most tightly bound of all (nickel-62 edges out iron-56 by a hair), so they sit at the bottom of the energy valley.

Checkpoint 43.2

Look at Figure 43.1. Explain, in terms of binding energy per nucleon, why *both* the fusion of very light nuclei and the fission of very heavy nuclei release energy, while neither process would release energy if it started from iron.

Answer

Example 43.2: Binding energy of mirror nuclei

Calculate the binding energy per nucleon for $^{23}_{11}\text{Na}$ and $^{23}_{12}\text{Mg}$, and account for the difference. Use atomic masses $m(^1\text{H}) = 1.007825\ \text{u}$, $m_n = 1.008665\ \text{u}$, $m(^{23}\text{Na}) = 22.989769\ \text{u}$, and $m(^{23}\text{Mg}) = 22.994124\ \text{u}$.

Solution

For $^{23}_{11}\text{Na}$ ($Z = 11$, $N = 12$), the mass of the separate parts is

$$\sum m = 11(1.007825) + 12(1.008665) = 23.190055\,\text{u}, \tag{43.2.1}$$

so the mass defect is $\sum m - M = 23.190055 - 22.989769 = 0.200286\ \text{u}$, and

$$E_b = (0.200286)(931.494) = 186.6\,\text{MeV}, \qquad E_{bn} = \frac{186.6}{23} = \boxed{8.11\,\text{MeV}}. \tag{43.2.2}$$

For $^{23}_{12}\text{Mg}$ ($Z = 12$, $N = 11$),

$$\sum m = 12(1.007825) + 11(1.008665) = 23.189215\,\text{u}, \tag{43.2.3}$$

$\sum m - M = 0.195091\ \text{u}$, so $E_b = 181.7\ \text{MeV}$ and

$$E_{bn} = \frac{181.7}{23} = \boxed{7.90\,\text{MeV}}. \tag{43.2.4}$$

Sodium is bound by about $0.21\ \text{MeV}$ per nucleon more than magnesium. The two nuclei have the same number of nucleons and nearly identical strong-force binding, but ^{23}Mg has one extra proton and one fewer neutron. The extra proton adds electrical repulsion, which *reduces* the binding, so the more proton-rich mirror nucleus is the less tightly bound of the pair.

Answers to Checkpoints

Checkpoint 43.2. Energy is released whenever nucleons end up more tightly bound, at a higher binding energy per nucleon, than they started. The curve rises steeply from the lightest nuclei to a peak near iron and then falls gently toward the heaviest. Fusing two light nuclei moves them up the rising side, and splitting a heavy nucleus moves its fragments up the falling side toward the peak: both raise the binding energy per nucleon and release energy. Iron sits at the top, so fusing it or splitting it would move the products *down* the curve; either process would absorb energy rather than release it.

This page titled 43.2: Nuclear Binding Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.3: Radioactivity

A nucleus that is not stable will sooner or later transform itself, emitting particles or radiation in the process. This spontaneous emission is **radioactivity**, and the transformation is **radioactive decay**. We cannot predict when any single nucleus will decay (it is a fundamentally random quantum event), but for a large sample the statistics are precise. The number of nuclei that decay in a short time is proportional to the number present, and this leads to exponential decay:

Key Equation: Radioactive decay law

$$N(t) = N_0\, e^{-\lambda t}, \tag{43.8}$$

where N_0 is the number of radioactive nuclei present at $t = 0$, $N(t)$ the number remaining at time t, and λ the **decay constant** (or disintegration constant), characteristic of the particular nuclide. The **activity** R is the rate of decays, the number of disintegrations per second:

$$R = \left|\frac{dN}{dt}\right| = \lambda N = \lambda N_0\, e^{-\lambda t} = R_0\, e^{-\lambda t}. \tag{43.9}$$

Activity is measured in **becquerels**, where $1\ \text{Bq} = 1$ decay per second. An older but still common unit is the **curie**: $1\ \text{Ci} \equiv 3.7 \times 10^{10}\ \text{Bq}$.

It is often more intuitive to speak not of the decay constant but of the **half-life** $T_{1/2}$: the time for half of any sample to decay. Setting $N = N_0/2$ in Equation (43.8) gives

Key Equation: Half-life

$$T_{1/2} = \frac{\ln 2}{\lambda} \approx \frac{0.693}{\lambda}. \tag{43.10}$$

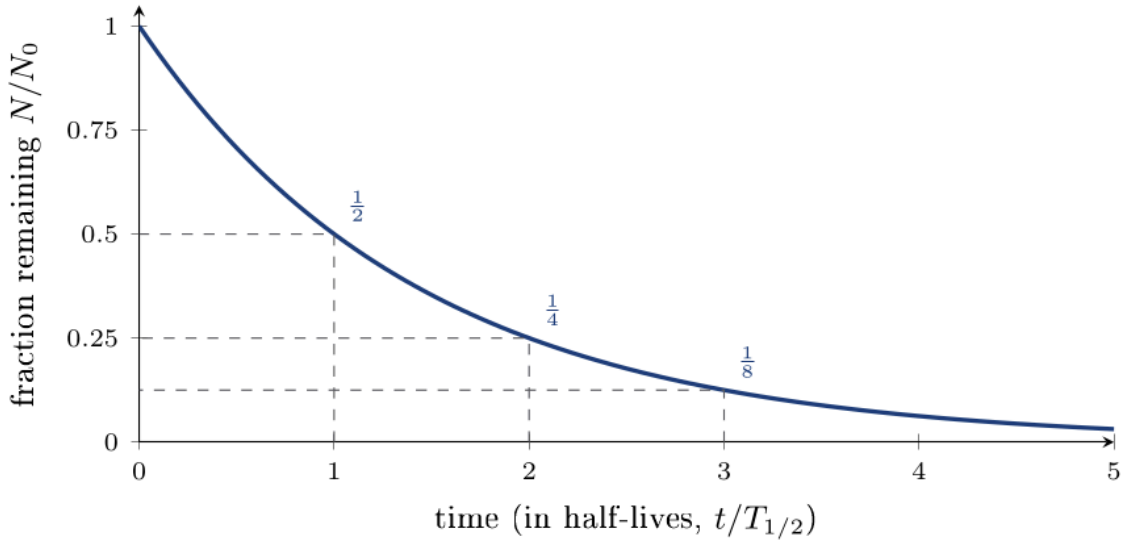


Figure 43.2. Exponential radioactive decay. After each half-life the number of surviving nuclei (and the activity, which is proportional to it) falls by a factor of two: to $\frac{1}{2}$, then $\frac{1}{4}$, then $\frac{1}{8}$ of the original. The curve never quite reaches zero, which is why a fixed half-life describes the decay no matter how much sample remains.

•Figure description

A graph of the surviving fraction N over N nought against time measured in half-lives, from 0 to 5. The curve starts at 1 and falls as a smooth exponential, never becoming straight and never quite reaching zero. Dashed guide lines mark the successive halvings: one half at one half-life, one quarter at two, one eighth at three. Each half-life removes half of whatever remains, regardless of how much that is.

After n half-lives have elapsed (so $t = n\ T_{1/2}$), the surviving fraction is simply

$$N = N_0 \left(\tfrac{1}{2}\right)^n, \tag{43.11}$$

as shown in Figure 43.2: half remain after one half-life, a quarter after two, an eighth after three, and so on.

Example 43.3: Activity of a carbon-11 sample

At $t = 0$ a sample contains $3.50\ \text{mg}$ of pure $^{11}_{6}\text{C}$, which has a half-life of $20.4\ \text{min}$. **(a)** Determine the number N_0 of nuclei present. **(b)** Find the activity initially and after $8.00\ \text{h}$.

Solution

(a) One mole of ^{11}C has a mass of about $11.0\ \text{g}$, so

$$N_0 = \frac{3.50 \times 10^{-3}\,\text{g}}{11.0\,\text{g/mol}}\,(6.022 \times 10^{23}\,\text{mol}^{-1}) = \boxed{1.92 \times 10^{20}\ \text{nuclei}}. \tag{43.3.1}$$

(b) The decay constant is

$$\lambda = \frac{\ln 2}{T_{1/2}} = \frac{0.693}{(20.4)(60)\,\text{s}} = 5.66 \times 10^{-4}\,\text{s}^{-1}. \tag{43.3.2}$$

The initial activity is

$$R_0 = \lambda N_0 = (5.66 \times 10^{-4})(1.92 \times 10^{20}) = \boxed{1.08 \times 10^{17}\,\text{Bq}}. \tag{43.3.3}$$

After $8.00\ \text{h} = 2.88 \times 10^4\ \text{s}$, we have $\lambda t = (5.66 \times 10^{-4})(2.88 \times 10^4) = 16.3$, so

$$R = R_0\, e^{-\lambda t} = (1.08 \times 10^{17})\, e^{-16.3} = \boxed{9.0 \times 10^{9}\,\text{Bq}}. \tag{43.3.4}$$

In eight hours (about 23.5 half-lives) the activity has fallen by a factor of $e^{16.3} \approx 1.2 \times 10^7$.

Radioactive Dating

Because each radioactive isotope decays with a fixed half-life, a decaying sample is a built-in clock. The best-known application is **radiocarbon dating**. Cosmic rays continually produce ^{14}C in the upper atmosphere, where it mixes in and is taken up by living things along with ordinary carbon, holding the ^{14}C fraction of a living organism constant. When the organism dies it stops taking in carbon, and its ^{14}C, with a half-life of $5730\ \text{yr}$, begins to decay away. By comparing the ^{14}C activity of an old sample to that of a living one, we can find how long ago it died. Solving Equation (43.9) for the elapsed time,

$$t = \frac{1}{\lambda}\ln\frac{R_0}{R} = \frac{T_{1/2}}{\ln 2}\ln\frac{R_0}{R}. \tag{43.12}$$

Example 43.4: Dating ancient charcoal

A piece of charcoal containing $25.0\ \text{g}$ of carbon is found in ancient ruins. Its measured ^{14}C activity is 250 decays/min. Living wood has a specific activity of about 15.0 decays/min per gram of carbon. How long ago did the tree die?

Solution

A living $25.0\ \text{g}$ sample would have an activity of

$$R_0 = (15.0\ \text{min}^{-1}\text{g}^{-1})(25.0\,\text{g}) = 375\ \text{decays/min}. \tag{43.3.5}$$

With the present activity $R = 250$ decays/min and $T_{1/2} = 5730\ \text{yr}$,

$$t = \frac{T_{1/2}}{\ln 2} \ln \frac{R_0}{R} = \frac{5730}{0.693} \ln \frac{375}{250} = (8266)(0.405) = \boxed{3.4 \times 10^3 \text{ yr}}. \quad (43.3.6)$$

The tree died roughly 3400 years ago. (Note that radiocarbon dating becomes unreliable beyond about ten half-lives, $\sim 50{,}000$ years, once too little ^{14}C remains to measure.)

Checkpoint 43.3

A radioactive isotope has a half-life of 5 days. (a) What fraction of a sample remains after 15 days? (b) One particular nucleus has already survived 10 days. Is it more likely to decay in the next 5 days than a freshly made nucleus of the same isotope?

Answer

Answers to Checkpoints

Checkpoint 43.3. (a) 15 days is three half-lives, so the fraction left is $(\frac{1}{2})^3 = \frac{1}{8}$. (b) No. Radioactive decay has no memory: each nucleus has the same probability of decaying in the next 5 days, one half, however long it has already survived. The half-life describes the chance of decay per unit time, not an aging process.

This page titled 43.3: Radioactivity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.4: Nuclear Reactions

When a nucleus is struck by a particle and transforms into different nuclei, we have a **nuclear reaction**. A general reaction in which a projectile a strikes a target X, producing b and Y, is written $a + X \to b + Y$. Two quantities are always conserved: the total number of nucleons (the sum of the mass numbers A) and the total charge (the sum of the atomic numbers Z). Balancing these is how we identify an unknown product.

The energy released or absorbed is the **reaction energy** or Q **value**, given by the change in rest mass:

Key Equation: Reaction energy (Q value)

$$Q = (M_a + M_X - M_b - M_Y)\,c^2. \tag{43.13}$$

If $Q > 0$, mass is converted to kinetic energy and the reaction is **exothermic** (energy is released). If $Q < 0$, the reaction is **endothermic** and cannot proceed unless the projectile brings in at least that much kinetic energy; the minimum required is the **threshold energy**. The decays we study next (alpha and beta decay) are simply spontaneous reactions with a single parent on the left, and their Q value is found the same way.

Example 43.5: Identifying a reaction product and its energy

Find the unknown product X and the reaction energy of

$$^{1}_{1}\mathrm{H} + {}^{19}_{9}\mathrm{F} \to {}^{16}_{8}\mathrm{O} + X. \tag{43.4.1}$$

Use atomic masses $m(^{1}\mathrm{H}) = 1.007825\ \mathrm{u}$, $m(^{19}\mathrm{F}) = 18.998403\ \mathrm{u}$, $m(^{16}\mathrm{O}) = 15.994915\ \mathrm{u}$, and $m(^{4}\mathrm{He}) = 4.002603\ \mathrm{u}$.

Solution

Conserve nucleons and charge. For mass number: $1 + 19 = 16 + A_X$, so $A_X = 4$. For charge: $1 + 9 = 8 + Z_X$, so $Z_X = 2$. The product is $^{4}_{2}\mathrm{He}$, an alpha particle:

$$^{1}_{1}\mathrm{H} + {}^{19}_{9}\mathrm{F} \to {}^{16}_{8}\mathrm{O} + {}^{4}_{2}\mathrm{He}. \tag{43.4.2}$$

The reaction energy is

$$Q = \left[(1.007825 + 18.998403) - (15.994915 + 4.002603)\right](931.494)\,\mathrm{MeV} \tag{43.4.3}$$

$$= (0.008710)(931.494) = \boxed{+8.11\,\mathrm{MeV}}. \tag{43.4.4}$$

Since $Q > 0$, the reaction is exothermic: it releases $8.11\ \mathrm{MeV}$.

Alpha Decay

In **alpha decay**, a heavy nucleus ejects an **alpha particle**: a tightly bound clump of two protons and two neutrons, identical to a $^{4}_{2}\mathrm{He}$ nucleus, with mass about $4\ \mathrm{u}$ and charge $+2e$. The original nucleus (the **parent**) thereby loses two protons and two neutrons, becoming a lighter **daughter** nucleus:

Key Equation: Alpha decay

$$^{A}_{Z}\mathrm{X} \;\to\; {}^{A-4}_{Z-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He}. \tag{43.14}$$

For example, the americium in household smoke detectors decays this way:

$$^{241}_{95}\mathrm{Am} \to {}^{237}_{93}\mathrm{Np} + {}^{4}_{2}\mathrm{He}. \tag{43.4.5}$$

About 99% of the helium on Earth comes from the alpha decay of underground uranium and thorium. Because alpha particles are heavy and doubly charged, they interact strongly with matter and lose their energy almost at once: a few centimeters of air, a sheet of paper, or even the dead outer layer of your skin will stop them. They are therefore relatively harmless externally, but dangerous if an alpha emitter is inhaled or ingested, where it can deposit all its energy directly into living tissue.

Example 43.6: The energy of an alpha decay

A ${}^{226}_{88}\mathrm{Ra}$ nucleus undergoes alpha decay to ${}^{222}_{86}\mathrm{Rn}$. **(a)** Find the Q value, using masses $226.025410\ \mathrm{u}$ for ${}^{226}\mathrm{Ra}$, $222.017578\ \mathrm{u}$ for ${}^{222}\mathrm{Rn}$, and $4.002603\ \mathrm{u}$ for ${}^{4}\mathrm{He}$. **(b)** What is the kinetic energy of the emitted alpha particle?

Solution

(a) The decay is ${}^{226}_{88}\mathrm{Ra} \to {}^{222}_{86}\mathrm{Rn} + {}^{4}_{2}\mathrm{He}$, so

$$Q = (226.025410 - 222.017578 - 4.002603)(931.494)\,\mathrm{MeV} \tag{43.4.6}$$

$$= (0.005229)(931.494) = \boxed{4.87\,\mathrm{MeV}}. \tag{43.4.7}$$

(b) The released energy is shared between the alpha particle and the recoiling daughter, which must carry equal and opposite momentum. Because the alpha is far lighter, it takes almost all the energy; momentum conservation gives

$$K_\alpha = Q\,\frac{M_{\mathrm{Rn}}}{M_{\mathrm{Rn}} + M_\alpha} \approx Q\,\frac{222}{226} = (4.87)(0.982) = \boxed{4.78\,\mathrm{MeV}}. \tag{43.4.8}$$

The daughter nucleus recoils with the small remaining $0.09\ \mathrm{MeV}$.

Beta Decay

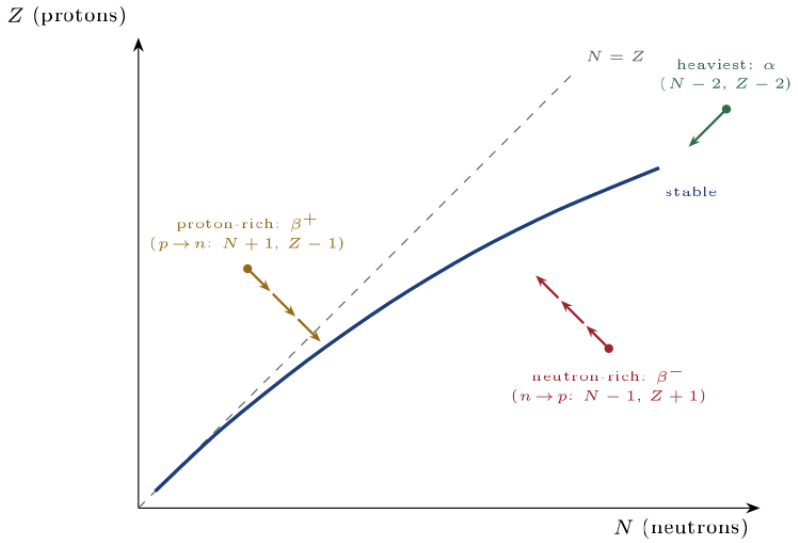


Figure 43.3. The valley of stability. Stable nuclei (blue band) cluster near $N = Z$ for light elements but bend toward neutron excess ($N > Z$) for heavy ones. Neutron-rich nuclei below the band decay by β^- (each step: $N - 1$, $Z + 1$) and proton-rich nuclei above it decay by β^+ ($N + 1$, $Z - 1$), each step nudging the nucleus toward the stable band; the heaviest nuclei decay by alpha emission ($N - 2$, $Z - 2$).

•Figure description

A chart of proton number Z against neutron number N. A dashed diagonal marks the line N equals Z. A thick blue band, the stable nuclei, follows that diagonal closely for light nuclei but bends progressively below it as the nuclei get heavier, so heavy stable nuclei have more neutrons than protons. The region below and to the right of the band is labeled neutron-rich, decaying by beta-minus as a neutron turns into a proton; the region above and to the left is labeled proton-rich, decaying by beta-plus as a proton turns into a neutron. Each decay moves a nucleus toward the band, shown by short arrows: a red dot below the band takes beta-minus steps diagonally up and to the left, a gold dot above the band takes beta-plus steps diagonally down and to the right, and a green dot beyond the far end of the band, the heaviest nuclei, takes an alpha step down and to the left by two in each direction.

Beta decay converts a neutron into a proton or vice versa, emitting a fast electron or positron (the **beta particle**) in the process. There are two types. In β^- **decay**, a neutron turns into a proton, emitting an electron and an (anti)neutrino:

$$n \to p + e^- + \bar{\nu}_e. \tag{43.15}$$

In β^+ **decay**, a proton turns into a neutron, emitting a positron and a neutrino. (This happens only inside a nucleus, which can supply the energy; a free proton is lighter than a neutron and cannot decay this way.)

$$p \to n + e^+ + \nu_e. \tag{43.16}$$

The nearly massless neutrino (ν_e) and antineutrino ($\bar{\nu}_e$) are emitted to conserve energy, momentum, and other quantities. In either case the mass number A is unchanged, while Z increases by one (β^-) or decreases by one (β^+). Examples include

$$^{14}_{6}\mathrm{C} \to {}^{14}_{7}\mathrm{N} + e^- + \bar{\nu}_e \quad (\beta^-), \qquad {}^{23}_{12}\mathrm{Mg} \to {}^{23}_{11}\mathrm{Na} + e^+ + \nu_e \quad (\beta^+). \tag{43.4.9}$$

Whether a nucleus decays by β^-, β^+, or not at all depends on its balance of neutrons and protons. Plotting the known nuclides on a chart of proton number Z versus neutron number N, the stable ones form a narrow band called the **valley of stability** (Figure 43.3). Light stable nuclei lie close to the line $N = Z$, but heavier ones need ever more neutrons (whose strong-force attraction dilutes the protons' repulsion) and the band bends toward $N > Z$. A nucleus lying *below* the valley has too many neutrons and decays by β^- (converting a neutron to a proton, which moves it one step up and one step left, toward the band); one lying *above* the valley is proton-rich and decays by β^+ (converting a proton to a neutron, one step down and one step right). The heaviest nuclei, beyond the end of the band, tend to shed whole alpha particles, which removes two protons and two neutrons at once.

Gamma decay. Alpha and beta decay often leave the daughter nucleus in an excited state. Just as an excited atom emits a photon when an electron drops to a lower level, an excited nucleus drops to its ground state by emitting a high-energy photon, a **gamma ray** (γ), with energy typically from tens of keV to a few MeV. Neither Z nor A changes. A few excited nuclei live long enough to be used on their own; technetium-99m (the "m" marks this long-lived, *metastable* excited state) emits a $140\ \mathrm{keV}$ gamma ray and is the workhorse of medical imaging.

The three kinds of radiation differ enormously in how far they travel through matter (Figure 43.4). Alpha particles, heavy and doubly charged, are stopped by a sheet of paper. Beta particles go farther but are stopped by a few millimeters of aluminum. Gamma rays, uncharged and massless, are the most penetrating: even several centimeters of lead only weaken them.

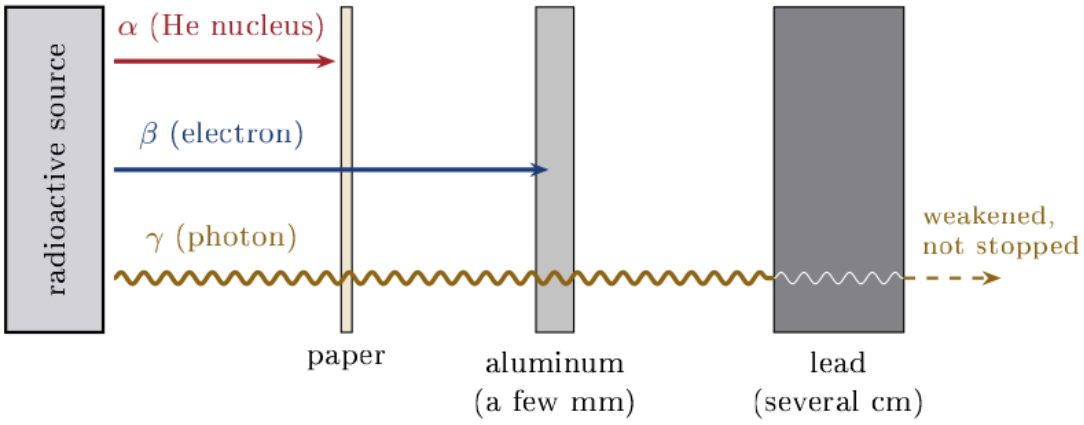


Figure 43.4. Penetrating power. Alpha particles are stopped by a sheet of paper and beta particles by a few millimeters of aluminum, but gamma rays are only weakened, never completely stopped, by several centimeters of lead.

•Figure description

The penetrating power of alpha, beta and gamma radiation. A box at the left labeled radioactive source emits three rays to the right toward three barriers: a thin sheet of paper, a few millimeters of aluminum, and a thick block of lead several centimeters wide. The red alpha arrow, a helium nucleus, stops at the paper. The blue beta arrow, an electron, passes through the paper and stops in the aluminum. The gold wavy gamma ray, a photon, passes through the paper and the aluminum, continues through the lead, and emerges on the far side as a dashed arrow labeled weakened, not stopped.

Example 43.7: Energy released in tritium decay

Tritium, $^{3}_{1}\mathrm{H}$, undergoes β^- decay to $^{3}_{2}\mathrm{He}$:

$$^{3}_{1}\mathrm{H} \rightarrow {}^{3}_{2}\mathrm{He} + e^- + \bar{\nu}_e. \tag{43.4.10}$$

Determine the total energy released. Use atomic masses $m(^3\mathrm{H}) = 3.016049\ \mathrm{u}$ and $m(^3\mathrm{He}) = 3.016029\ \mathrm{u}$.

Solution

When atomic (rather than nuclear) masses are used, the mass of the emitted β^- electron is automatically accounted for, because the daughter atom $^3\mathrm{He}$ already includes one more atomic electron than the parent. The Q value is therefore simply

$$Q = \left[m(^3\mathrm{H}) - m(^3\mathrm{He})\right]c^2 = (3.016049 - 3.016029)(931.494)\,\mathrm{MeV} \tag{43.4.11}$$

$$= (0.0000200)(931.494) = 0.0186\,\mathrm{MeV} = \boxed{18.6\,\mathrm{keV}}. \tag{43.4.12}$$

This tiny energy release, shared between the electron and the antineutrino, is precisely why tritium decay is used in modern experiments that attempt to measure the mass of the neutrino.

Checkpoint 43.4

A nucleus has many more neutrons than its stable isotopes do, placing it below the valley of stability. Will it most likely decay by β^- or β^+? Explain which way the decay moves it on the Z-versus-N chart.

Answer

Nuclear Fission

Nuclear fission is the splitting of a heavy nucleus into two lighter ones, with the release of a large amount of energy. The reason energy is released is exactly the binding-energy curve of Figure 43.1: a heavy nucleus like uranium sits on the gentle right-hand slope at about $7.6\ \mathrm{MeV}$ per nucleon, but its mid-mass fragments lie higher, near $8.5\ \mathrm{MeV}$ per nucleon. The nucleons end up more tightly bound, and the difference emerges as energy.

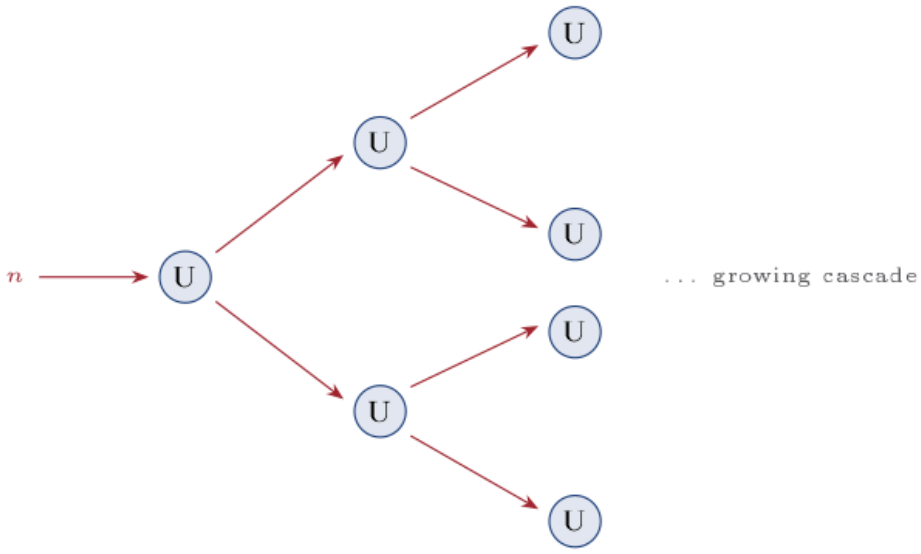


Figure 43.5. A fission chain reaction. A single neutron triggers one fission, which releases two or more neutrons; each of those can trigger another fission, and the number of events multiplies from one generation to the next. Controlling this multiplication (keeping it at exactly one new fission per event) is what a reactor must do.

•Figure description

A fission chain reaction spreading left to right through three generations. A single red neutron arrow at the left strikes one uranium nucleus, drawn as a circle marked U. Two red neutron arrows leave it, striking two more uranium nuclei. Four arrows leave those, striking four more. The number of nuclei doubles at each step (one, then two, then four), and a note at the right reads growing cascade. Keeping this multiplication at exactly one is what a reactor must do.

A typical fission begins when a $^{235}\mathrm{U}$ nucleus absorbs a slow (“thermal”) neutron, briefly becoming highly unstable $^{236}\mathrm{U}$, which then splits (Figure 43.6):

$$^{235}_{92}\mathrm{U} + n \rightarrow ^{236}_{92}\mathrm{U} \rightarrow ^{140}_{54}\mathrm{Xe} + ^{94}_{38}\mathrm{Sr} + 2n. \tag{43.17}$$

The absorbed neutron’s binding energy leaves the $^{236}\mathrm{U}$ nucleus excited, and it vibrates like a liquid drop. If it stretches far enough, the short-range strong force can no longer hold the two ends together, while the long-range electric repulsion between them keeps pushing, and the nucleus pulls apart.

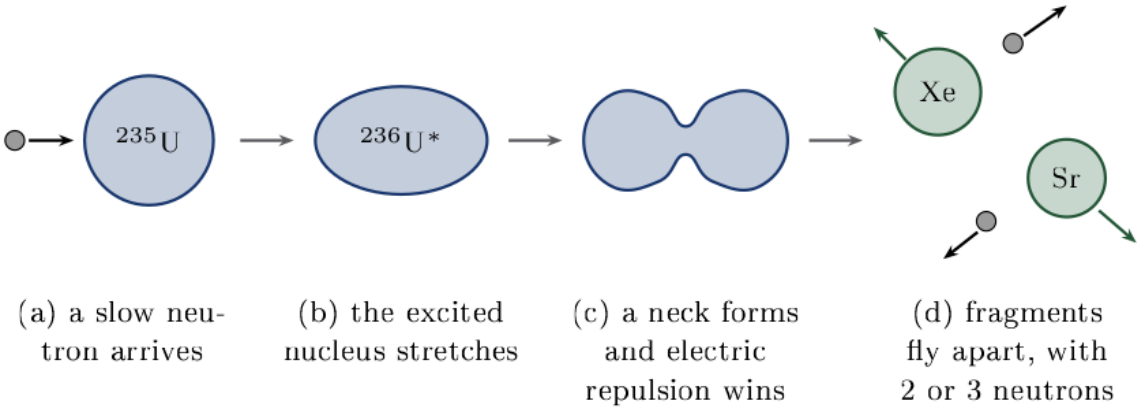


Figure 43.6. The stages of fission, pictured with the liquid-drop model of the nucleus. Once the excited ^{236}U stretches into a neck, the long-range electric repulsion between the two halves overwhelms the short-range strong force, and the nucleus splits into two fragments plus a few neutrons.

•Figure description

Four stages of a fission event, left to right. Panel a: a small gray neutron with an arrow approaches a round blue nucleus labeled uranium-235, with the caption a slow neutron arrives. Panel b: an elongated, egg-shaped nucleus labeled uranium-236 star, with the caption the excited nucleus stretches. Panel c: a peanut-shaped nucleus pinched in the middle, with the caption a neck forms and electric repulsion wins. Panel d: two green fragments labeled Xe and Sr flying apart in opposite directions, with two small gray neutrons also flying off, with the caption fragments fly apart, with 2 or 3 neutrons.

The fragments are not fixed (many different pairs are possible) but they are always neutron-rich (having come from a neutron-heavy parent), so they go on to reach stability through chains of β^- decays. The two or three neutrons released are the key to a self-sustaining **chain reaction**: each can trigger another fission, which releases more neutrons, and so on (Figure 43.5). If on average more than one neutron from each fission goes on to cause another, the reaction grows; if fewer than one, it dies out. A reactor holds the average at exactly one, while a weapon lets the reaction grow as fast as possible.

The energy per fission can be estimated directly from the binding-energy curve. Since the number of nucleons is conserved, the released energy is

$$Q = A\left[(E_{bn})_{\text{final}} - (E_{bn})_{\text{initial}}\right]. \tag{43.18}$$

For uranium, with $A \approx 238$ and a gain of roughly $0.9\ \text{MeV}$ per nucleon, this gives about $200\ \text{MeV}$ per fission: tens of millions of times the few electron-volts released in a chemical reaction such as burning.

Example 43.8: Energy from a kilogram of uranium

Calculate the energy released when $1.00\ \text{kg}$ of ^{235}U undergoes fission, taking the energy per fission event to be $Q = 208\ \text{MeV}$.

Solution

The number of ^{235}U nuclei in $1.00\ \text{kg}$ is

$$N = \frac{1000\,\text{g}}{235\,\text{g/mol}}(6.022\times10^{23}\,\text{mol}^{-1}) = 2.56\times10^{24}. \tag{43.4.13}$$

The total energy is

$$E = NQ = (2.56\times10^{24})(208\,\text{MeV}) = 5.33\times10^{26}\,\text{MeV}. \tag{43.4.14}$$

Converting to joules with $1\ \text{MeV} = 1.602\times10^{-13}\ \text{J}$,

$$E = (5.33\times10^{26})(1.602\times10^{-13}\,\text{J}) = \boxed{8.5\times10^{13}\,\text{J}}. \tag{43.4.15}$$

This is comparable to burning several thousand tonnes of coal: the energy density of nuclear fuel is staggering.

Nuclear Power Plants

Every commercial nuclear power plant operating today runs on controlled fission. The aim is to sustain a chain reaction at a steady rate (exactly one new fission per fission, a state called **criticality**) rather than letting it grow explosively or die out. A reactor core contains several key elements. The **fuel**, usually uranium enriched to a few percent $^{235}\mathrm{U}$, is formed into rods. A **moderator** (often water) slows the fast fission neutrons to thermal speeds, where they are far more likely to cause further fission. **Control rods** of neutron-absorbing material (such as cadmium or boron) are inserted or withdrawn to tune the reaction rate, soaking up excess neutrons. A **coolant** carries the heat away to boil water into steam, which drives turbines and generators just as in any thermal power station. The nuclear core, in the end, is simply a very compact and very long-lived heat source.

Example 43.9: Inside a power reactor

A power station is driven by a pressurized-water reactor whose core produces $3400\ \mathrm{MW}$ of thermal power, of which $1100\ \mathrm{MW}$ is delivered as electricity. **(a)** What is the station's efficiency? **(b)** At what rate R do fission events occur, taking $208\ \mathrm{MeV}$ per fission? **(c)** At what rate is mass being converted to energy?

Solution

(a) Efficiency is useful output over input,

$$e = \frac{1100\,\mathrm{MW}}{3400\,\mathrm{MW}} = \boxed{0.324} = 32.4\%. \tag{43.4.16}$$

(b) Each fission delivers $208\ \mathrm{MeV} = (208)(1.602\times10^{-13}\ \mathrm{J}) = 3.33\times10^{-11}\ \mathrm{J}$ of thermal energy, so to produce $3400\ \mathrm{MW} = 3.40\times10^{9}\ \mathrm{W}$,

$$R = \frac{3.40\times10^{9}\,\mathrm{W}}{3.33\times10^{-11}\,\mathrm{J}} = \boxed{1.0\times10^{20}\ \mathrm{fissions/s}}. \tag{43.4.17}$$

(c) The mass converted to energy follows from $P = (dm/dt)c^2$:

$$\frac{dm}{dt} = \frac{P}{c^2} = \frac{3.40\times10^{9}\,\mathrm{W}}{(3.00\times10^{8}\,\mathrm{m/s})^2} = \boxed{3.8\times10^{-8}\,\mathrm{kg/s}}, \tag{43.4.18}$$

only about $3\ \mathrm{g}$ per day, yet that tiny mass loss is the entire source of the plant's enormous output.

Nuclear Fusion

At the opposite end of the binding-energy curve, energy can be released by joining light nuclei together rather than splitting heavy ones. This is **nuclear fusion**, and per kilogram of fuel it releases even more energy than fission, because the left-hand slope of the binding-energy curve is so steep. It is also the process that powers the stars.

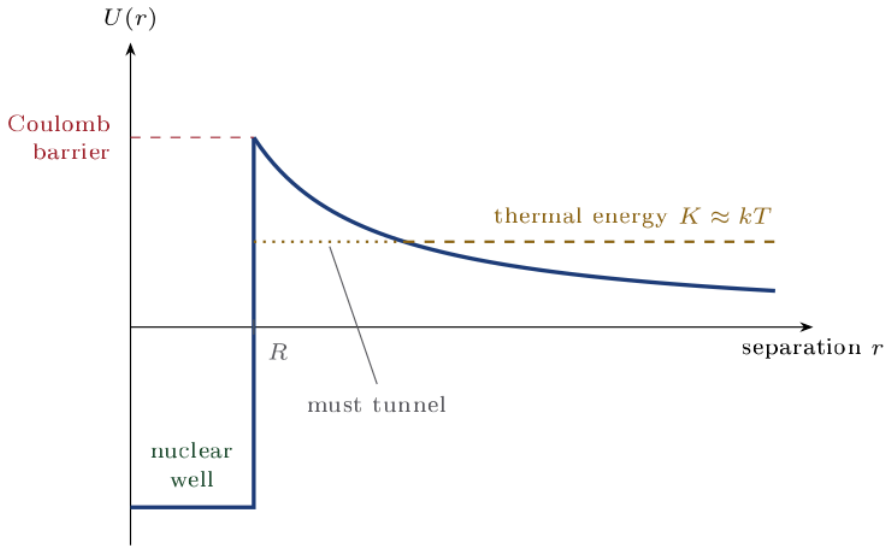


Figure 43.7. Potential energy of two approaching nuclei. As they near each other the Coulomb repulsion drives $U(r)$ up like $1/r$, forming a barrier; only once they are within about $r = R$ does the strong force take over and pull them into the deep attractive well, fusing them. Typical thermal energies (gold) fall far below the barrier peak, which is why fusion requires enormous temperatures.

•Figure description

A graph of potential energy U against the separation r of two approaching nuclei. Coming in from large separation on the right, the curve rises like one over r as the nuclei repel electrically, reaching a peak marked by a dashed red line labeled Coulomb barrier at a separation r equals capital R. Inside that radius the curve plunges steeply through zero into a deep negative well labeled nuclear well, where the short-range strong force takes over and binds the nuclei. A gold dashed line drawn well below the top of the barrier marks typical thermal energy k T; it meets the curve well outside R, and a dotted continuation labeled must tunnel shows the stretch of barrier a nucleus with that energy could cross only by quantum tunneling.

Fusion is far harder to achieve than fission, however, because both nuclei are positively charged and repel each other fiercely. To get them close enough for the short-range strong force to take hold and bind them, they must overcome the **Coulomb barrier**, the hump in potential energy created by their electrical repulsion (Figure 43.7). For two light nuclei this barrier is of order a few hundred keV. Classically there is no way around it. Quantum mechanically the nuclei need not climb all the way to the top, because they can tunnel through the upper part of the barrier (Section 40.4), but they must still be slammed together hard enough to get close.

The practical way to supply that energy in bulk is heat. The average thermal kinetic energy of a particle is $K \approx kT$, and at room temperature this is only about $0.03\ \text{eV}$: ten million times too small. Only at temperatures of tens of millions of kelvin do nuclei move fast enough to fuse on their own; fusion driven by such extreme heat is called **thermonuclear fusion**. Even in the Sun's core, at $1.5 \times 10^7\ \text{K}$, the typical thermal energy kT is only about $1\ \text{keV}$, hundreds of times below the barrier. Fusion happens there only because the fastest few nuclei in the thermal distribution tunnel through the barrier, which is one reason the Sun burns its fuel so slowly. Containing matter that hot is the central engineering challenge: the two leading approaches are **magnetic confinement**, which traps a hot plasma in a doughnut-shaped magnetic field (the tokamak, such as ITER, under construction in France), and **inertial confinement**, which crushes and heats a tiny fuel pellet with intense laser beams (as at the National Ignition Facility in California). Fusion's appeal is great (abundant fuel, no long-lived radioactive waste, no possibility of a meltdown). In 2022 the National Ignition Facility became the first to get more fusion energy out of a pellet than the laser light delivered to it, but no fusion device has yet produced more energy than the whole facility consumes, and reaching that point remains one of the hardest problems in applied physics.

Example 43.10: Deuterium–tritium fusion

A leading candidate fusion reaction combines deuterium and tritium:

$$^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + n. \tag{43.4.19}$$

Find the energy released, using masses $m(^2\text{H}) = 2.014102\ \text{u}$, $m(^3\text{H}) = 3.016049\ \text{u}$, $m(^4\text{He}) = 4.002603\ \text{u}$, and $m_n = 1.008665\ \text{u}$.

Solution

The Q value is the mass defect times c^2:

$$Q = \big[(2.014102 + 3.016049) - (4.002603 + 1.008665)\big](931.494)\,\text{MeV} \tag{43.4.20}$$

$$= (0.018883)(931.494) = \boxed{17.6\,\text{MeV}}. \tag{43.4.21}$$

This single reaction releases $17.6\ \text{MeV}$, and crucially, it does so from just five nucleons, so the energy *per nucleon* (about $3.5\ \text{MeV}$) far exceeds that of fission (about $0.9\ \text{MeV}$). That is why fusion fuel is so energy-dense.

In the Sun, fusion proceeds by the **proton–proton chain**, whose net effect is to combine four protons into one helium nucleus:

$$4\,{}^{1}\text{H} \rightarrow {}^{4}\text{He} + 2e^{+} + 2\nu_e + \text{energy}. \tag{43.4.22}$$

Each completed chain releases about $26.7\ \text{MeV}$, of which roughly $0.5\ \text{MeV}$ escapes immediately with the neutrinos, leaving $26.2\ \text{MeV}$ as the thermal energy that ultimately streams away as sunlight. The same fusion processes, occurring in stars and in their explosive deaths, are also where almost all the chemical elements heavier than hydrogen were forged: the carbon in your body and the iron in your blood were made inside stars.

Checkpoint 43.5

Per nucleon, fusion releases several times as much energy as fission, and its fuel is far more abundant. Yet we generate power commercially with fission and not (yet) with fusion. Referring to the Coulomb barrier of Figure 43.7, explain the basic physical reason fusion is so much harder to achieve.

Answer

Answers to Checkpoints

Checkpoint 43.4. β^- decay. A neutron-rich nucleus has too many neutrons for its protons. In β^- decay a neutron turns into a proton, emitting an electron and an antineutrino, so N drops by one and Z rises by one. That moves the nucleus diagonally up toward the valley of stability. (β^+ decay would do the opposite, and is the path for proton-rich nuclei.)

Checkpoint 43.5. In fusion, both nuclei are positively charged, so they repel each other electrically. The strong force that binds them acts only when they come within a few femtometers, and reaching that distance means climbing the Coulomb barrier of Figure 43.7. That takes enormous kinetic energies (temperatures of order $10^8\ \text{K}$) and holding such a hot plasma together long enough for many fusions is extremely difficult. Fission faces no such barrier: it is triggered by a neutron, which has no charge and can drift into a uranium nucleus at low speed.

This page titled 43.4: Nuclear Reactions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.S: Nuclear Physics (Summary)

Summary

- A nucleus has Z protons, N neutrons, and mass number $A = Z + N$, written ${}^{A}_{Z}\mathrm{X}$. Its radius is $r = aA^{1/3}$ with $a = 1.2\ \mathrm{fm}$, so all nuclei have nearly the same density. Masses are measured in u, with $c^2 = 931.494\ \mathrm{MeV/u}$. The short-range **strong force** binds the nucleons.
- **Binding energy**: $E_b = (\sum m - M)c^2$; **per nucleon** $E_{bn} = E_b/A$. The curve of E_{bn} versus A peaks near ${}^{56}\mathrm{Fe}$ ($\approx 8.8\ \mathrm{MeV}$); energy is released by moving toward this peak: fusing light nuclei or splitting heavy ones.
- **Radioactivity**: $N = N_0 e^{-\lambda t}$ and activity $R = \lambda N = R_0 e^{-\lambda t}$ (units: $1\ \mathrm{Bq} = 1$ decay/s, $1\ \mathrm{Ci} = 3.7 \times 10^{10}\ \mathrm{Bq}$). The half-life is $T_{1/2} = \ln 2/\lambda$, and after n half-lives $N = N_0(\frac{1}{2})^n$. Fixed half-lives make radioactive dating possible (e.g. ${}^{14}\mathrm{C}$, $T_{1/2} = 5730\ \mathrm{yr}$).
- **Nuclear reactions** conserve A and Z; the energy released is $Q = (M_a + M_X - M_b - M_Y)c^2$, with $Q > 0$ exothermic and $Q < 0$ endothermic (requiring a threshold energy).
- **Alpha decay**: ${}^{A}_{Z}\mathrm{X} \to {}^{A-4}_{Z-2}\mathrm{Y} + {}^{4}_{2}\mathrm{He}$; the light alpha carries off most of the Q value. **Beta decay**: β^- ($n \to p + e^- + \bar{\nu}_e$, Z up) and β^+ ($p \to n + e^+ + \nu_e$, Z down), driving nuclei toward the **valley of stability**.
- **Fission**: a heavy nucleus splits (e.g. ${}^{235}\mathrm{U} + n \to {}^{140}\mathrm{Xe} + {}^{94}\mathrm{Sr} + 2n$), releasing $\sim 200\ \mathrm{MeV}$; the freed neutrons sustain a **chain reaction**, controlled at criticality in reactors.
- **Fusion**: light nuclei join (e.g. ${}^{2}\mathrm{H} + {}^{3}\mathrm{H} \to {}^{4}\mathrm{He} + n$, $Q = 17.6\ \mathrm{MeV}$), releasing even more energy per nucleon but requiring enormous temperatures to overcome the **Coulomb barrier**. The Sun fuses hydrogen to helium via the proton–proton chain ($\approx 26.7\ \mathrm{MeV}$ per chain).

Key Equations

Mass number and nuclear radius

$$A = Z + N, \qquad r = aA^{1/3}, \quad a = 1.2\,\mathrm{fm} \tag{43.S.1}$$

Mass–energy conversion

$$c^2 = 931.494\,\mathrm{MeV/u} \tag{43.S.2}$$

Nuclear binding energy

$$E_b = \left(\sum m - M\right) c^2, \qquad E_{bn} = \frac{E_b}{A} \tag{43.S.3}$$

Radioactive decay law

$$N(t) = N_0\, e^{-\lambda t} \tag{43.S.4}$$

Activity

$$R = \left|\frac{\mathrm{d}N}{\mathrm{d}t}\right| = \lambda N = R_0\, e^{-\lambda t} \qquad (1\,\mathrm{Bq} = 1\ \mathrm{decay/s},\ 1\,\mathrm{Ci} = 3.7 \times 10^{10}\,\mathrm{Bq}) \tag{43.S.5}$$

Half-life

$$T_{1/2} = \frac{\ln 2}{\lambda} \approx \frac{0.693}{\lambda}, \qquad N = N_0\left(\tfrac{1}{2}\right)^n \text{ after } n \text{ half-lives} \tag{43.S.6}$$

Radioactive dating

$$t = \frac{1}{\lambda}\ln\frac{R_0}{R} = \frac{T_{1/2}}{\ln 2}\ln\frac{R_0}{R} \tag{43.S.7}$$

Reaction energy (Q value)

$$Q = (M_a + M_X - M_b - M_Y)\,c^2 \qquad (Q > 0 \text{ exothermic},\ Q < 0 \text{ endothermic}) \tag{43.S.8}$$

Alpha decay

$$^{A}_{Z}\mathrm{X} \ \rightarrow\ ^{A-4}_{Z-2}\mathrm{Y} + ^{4}_{2}\mathrm{He} \tag{43.S.9}$$

Beta decay

$$\begin{aligned} \beta^- &: \quad n \rightarrow p + e^- + \bar{\nu}_e \qquad (Z \text{ up by } 1) \\ \beta^+ &: \quad p \rightarrow n + e^+ + \nu_e \qquad (Z \text{ down by } 1) \end{aligned} \tag{43.S.10}$$

Energy released in fission

$$Q = A\left[(E_{bn})_{\text{final}} - (E_{bn})_{\text{initial}}\right] \tag{43.S.11}$$

This page titled 43.S: Nuclear Physics (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

43.E: Nuclear Physics (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Use these values: $c^2 = 931.494\ \text{MeV/u}$, $1\ \text{u} = 1.66 \times 10^{-27}\ \text{kg}$, $m(^1\text{H}) = 1.007825\ \text{u}$, $m_n = 1.008665\ \text{u}$, $N_A = 6.022 \times 10^{23}\ \text{mol}^{-1}$, $1\ \text{MeV} = 1.602 \times 10^{-13}\ \text{J}$, $1\ \text{Ci} = 3.7 \times 10^{10}\ \text{Bq}$, ^{14}C half-life $5730\ \text{yr}$, and $a = 1.2\ \text{fm}$.

Properties of Nuclei

43.1 ★ State the number of protons, neutrons, and nucleons in each of $^{12}_{6}\text{C}$, $^{56}_{26}\text{Fe}$, and $^{238}_{92}\text{U}$.

43.2 ★★ Estimate the radius of a ^{56}Fe nucleus, and find how many times larger it is than a single nucleon ($A = 1$).

43.3 ★★ *[Conceptual.]* Using $r = aA^{1/3}$, explain why all nuclei have approximately the same density, and state what that implies about how nuclear volume depends on A.

Nuclear Binding Energy

43.4 ★★ The deuteron $^{2}_{1}\text{H}$ has atomic mass $2.014102\ \text{u}$. Find its binding energy and its binding energy per nucleon.

43.5 ★★ The $^{4}_{2}\text{He}$ nucleus has atomic mass $4.002603\ \text{u}$. Find its total binding energy and compare its binding energy per nucleon to that of the deuteron.

43.6 ★★★ *[Real-world application: astrophysics — why stars stop at iron.]* The energy released when nuclei rearrange is

$$Q = (\text{total binding energy after}) - (\text{total binding energy before}). \tag{43.E.1}$$

Use these approximate binding energies per nucleon: ^4He, $7.07\ \text{MeV}$; ^{12}C, $7.68\ \text{MeV}$; ^{28}Si, $8.45\ \text{MeV}$; ^{56}Ni, $8.64\ \text{MeV}$; ^{56}Fe, $8.79\ \text{MeV}$; and a nucleus with $A = 112$, about $8.5\ \text{MeV}$.

a. Estimate Q for the “triple-alpha” process, three ^4He nuclei fusing into ^{12}C.
b. Estimate Q for two ^{28}Si nuclei fusing into ^{56}Ni, a late stage in massive stars.
c. Estimate Q for two ^{56}Fe nuclei fusing into a nucleus with $A = 112$.
d. A star’s core shines by fusing ever heavier nuclei, but once it is made of iron the fusion fire goes out. Use your answers to explain why.

Radioactivity

43.7 ★ A radioactive source has a half-life of $8.0\ \text{days}$. What fraction of the original sample remains after $24\ \text{days}$?

43.8 ★★ A sample of ^{32}P ($T_{1/2} = 14.3\ \text{days}$) has an initial activity of $6.0\ \text{mCi}$. Find its decay constant and its activity after $30\ \text{days}$.

43.9 ★★ A radioactive sample contains 4.0×10^{18} nuclei of an isotope whose half-life is $12\ \text{h}$. Find the decay constant and the initial activity in becquerels.

Radioactive Dating

43.10 ★★ *[Real-world application: archaeology.]* A wooden artifact shows a ^{14}C activity that is 0.25 times that of living wood. How old is the artifact?

43.11 ★★★ *[Real-world application: archaeology.]* A bone sample has a ^{14}C activity of 2.0 decays/min per gram of carbon, compared with 15.0 for living tissue. Estimate the age of the bone, and comment on whether radiocarbon dating is reliable at this age.

Nuclear Reactions

43.12 ★ Balance the reaction $^{14}_{7}\mathrm{N} + {}^{4}_{2}\mathrm{He} \to {}^{17}_{8}\mathrm{O} + X$ by identifying the particle X.

43.13 ★★ Find the reaction energy Q of $^{7}_{3}\mathrm{Li} + {}^{1}_{1}\mathrm{H} \to {}^{4}_{2}\mathrm{He} + {}^{4}_{2}\mathrm{He}$, using $m(^{7}\mathrm{Li}) = 7.016004\ \mathrm{u}$ and $m(^{4}\mathrm{He}) = 4.002603\ \mathrm{u}$. Is it exothermic or endothermic?

Alpha Decay

43.14 ★ A $^{238}_{92}\mathrm{U}$ nucleus undergoes alpha decay. Write the decay and identify the daughter nucleus.

43.15 ★★ $^{210}_{84}\mathrm{Po}$ alpha-decays to $^{206}_{82}\mathrm{Pb}$. Using $m(^{210}\mathrm{Po}) = 209.982874\ \mathrm{u}$, $m(^{206}\mathrm{Pb}) = 205.974465\ \mathrm{u}$, and $m(^{4}\mathrm{He}) = 4.002603\ \mathrm{u}$, find the Q value of the decay.

Beta Decay

43.16 ★ Complete each decay and state whether it is β^- or β^+: (a) $^{14}_{6}\mathrm{C} \to {}^{14}_{7}\mathrm{N} + \ldots$; (b) $^{22}_{11}\mathrm{Na} \to {}^{22}_{10}\mathrm{Ne} + \ldots$

43.17 ★★ *[Real-world application: medicine — PET scans.]* Carbon-11 is used in positron emission tomography (PET). It decays by β^+ emission to boron-11.

a. Write the decay.
b. Using *atomic* masses, the parent atom has Z electrons but the daughter atom needs only $Z - 1$, and a positron is created as well. Show that

$$Q = [m(\text{parent}) - m(\text{daughter}) - 2m_e]\,c^2. \tag{43.E.2}$$

c. Find Q using $m(^{11}\mathrm{C}) = 11.011434\ \mathrm{u}$, $m(^{11}\mathrm{B}) = 11.009305\ \mathrm{u}$, and $m_e = 0.000549\ \mathrm{u}$.
d. What would you get if you forgot the $2m_e$ term? Why does no such term appear in β^- decay?

Nuclear Fission

43.18 ★★ In a fission event, $^{235}_{92}\mathrm{U}$ absorbs a neutron and splits into $^{141}_{56}\mathrm{Ba}$, another fragment, and three neutrons. Identify the second fragment.

43.19 ★★★ Using the binding-energy curve, estimate the energy released when a $^{238}\mathrm{U}$ nucleus ($E_{bn} \approx 7.6\ \mathrm{MeV}$) splits into two equal fragments with $E_{bn} \approx 8.5\ \mathrm{MeV}$. Is the process exothermic?

Nuclear Power Plants

43.20 ★★ A reactor core generates $2500\ \mathrm{MW}$ of thermal power. Taking $200\ \mathrm{MeV}$ per fission, find the rate at which $^{235}\mathrm{U}$ nuclei undergo fission.

43.21 ★★★ *[Real-world application: energy.]* The reactor of the previous problem runs for one year. Estimate the mass of $^{235}\mathrm{U}$ consumed, assuming each fission uses one nucleus and the plant runs continuously.

Nuclear Fusion

43.22 ★★ Find the energy released in the fusion reaction $^{2}_{1}\mathrm{H} + {}^{2}_{1}\mathrm{H} \to {}^{3}_{2}\mathrm{He} + n$, using $m(^{2}\mathrm{H}) = 2.014102\ \mathrm{u}$, $m(^{3}\mathrm{He}) = 3.016029\ \mathrm{u}$, and $m_n = 1.008665\ \mathrm{u}$.

43.23 ★★★ *[Estimation.]* Two protons must come within about $2\ \mathrm{fm}$ of each other for the strong force to take hold. Use $ke^2 = 1.44\ \mathrm{MeV}\cdot\mathrm{fm}$ and Boltzmann's constant $k_B = 8.617 \times 10^{-5}\ \mathrm{eV/K}$.

a. Estimate the height of the Coulomb barrier, ke^2/r, at $r = 2\ \mathrm{fm}$.

b. At what temperature would the typical thermal energy $k_B T$ equal this barrier?
c. The core of the Sun is at about 1.5×10^7 K. By what factor does your answer to (b) exceed this? Explain how protons in the Sun manage to fuse anyway.

Bringing It Together — the Chapter Callback

43.24 ★★★ *[Real-world application: energy density — the chapter callback.]* Consider powering a single 100 W light bulb with the energy stored in 1.00 kg of fuel, by three different processes.

a. **Chemical:** burning coal releases about 3.3×10^7 J/kg. How long can 1 kg of coal keep the bulb lit?
b. **Fission:** from the worked example, 1 kg of $^{235}\mathrm{U}$ releases 8.5×10^{13} J. How long now?
c. **Fusion:** the complete fusion of 1 kg of hydrogen fuel releases roughly 6×10^{14} J. How long in this case?
d. Compare the three times and relate the enormous differences back to the curve of binding energy and $E = mc^2$: the theme with which this chapter began.

Challenge Problems

43.25 ★★★ *[Real-world application: medicine — scheduling a nuclear scan.]* The isotope technetium-99m is the workhorse of diagnostic nuclear medicine: it emits a convenient 140 keV gamma ray and has a half-life of 6.01 hr. A patient is scheduled for a bone scan at 9:00 a.m. and requires an administered activity of 740 MBq. The pharmacy prepares the dose at 6:00 a.m.

a. Find the decay constant, in hr^{-1} and in s^{-1}.
b. What activity must the syringe hold at 6:00 a.m. so that 740 MBq remains at 9:00 a.m.?
c. How many $^{99m}\mathrm{Tc}$ nuclei are present in the 740 MBq dose? What is their total mass, in grams? Comment on the result.
d. A six-hour half-life is short enough that the isotope cannot be shipped far, yet it is used in tens of millions of scans a year. Explain what makes such a short half-life desirable for a patient, and describe in one or two sentences how a $^{99}\mathrm{Mo}$ "generator" ($T_{1/2} = 66$ hr), which decays into $^{99m}\mathrm{Tc}$, solves the supply problem.

43.26 ★★★ *[Real-world application: geology — dating a zircon crystal.]* Uranium-238 decays through a long chain that terminates at stable lead-206, with an effective half-life of 4.468×10^9 yr. Zircon crystals incorporate uranium when they form but reject lead, so any $^{206}\mathrm{Pb}$ found inside one was made *in situ*. A grain is measured to contain 0.0725 atoms of $^{206}\mathrm{Pb}$ for every atom of $^{238}\mathrm{U}$.

a. Let N_0 be the original number of $^{238}\mathrm{U}$ atoms. Write $N_{\mathrm{U}}(t)$ and $N_{\mathrm{Pb}}(t)$, and show that

$$\frac{N_{\mathrm{Pb}}}{N_{\mathrm{U}}} = e^{\lambda t} - 1. \tag{43.E.3}$$

b. Find the decay constant and the age of the grain.
c. Carbon-14 dating ($T_{1/2} = 5730$ yr) is useless here. Show this quantitatively by finding how many $^{14}\mathrm{C}$ half-lives fit into the age you found in (b), and comment on what would be left of the sample. Then state the general rule relating a radioactive clock's half-life to the range of ages it can usefully measure, and explain why $^{238}\mathrm{U}$ is equally useless for dating a 3000 yr-old artifact.
d. The calculation in (a) assumed the grain contained *no* lead when it formed and has neither gained nor lost uranium or lead since. State how each assumption would bias the computed age if it were violated, and suggest how a geologist might test them.

43.27 ★★★ *[Real-world application: astrophysics — what powers the Sun.]* The proton–proton chain converts four hydrogen nuclei into one helium-4 nucleus. Using atomic masses (which include the electrons, so that the two emitted positrons and their annihilation are accounted for automatically), the net reaction is $4\,{}^1\mathrm{H} \to {}^4\mathrm{He}$, with $m({}^1\mathrm{H}) = 1.007825$ u and $m({}^4\mathrm{He}) = 4.002603$ u. The Sun's luminosity is $L_\odot = 3.85 \times 10^{26}$ W and its mass is 1.99×10^{30} kg.

a. Find the mass difference in atomic mass units and the energy released per reaction, in MeV.
b. What fraction of the initial mass is converted to energy?
c. How many such reactions occur in the Sun each second? How many kilograms of hydrogen are consumed per second?
d. The Sun's *mass loss* rate is $L_\odot / c^2$. Evaluate it, and reconcile it with your (much larger) answer for hydrogen consumed in (c). Then estimate how long the Sun could shine at this rate if it could burn 10% of its total mass as hydrogen, and compare with the accepted main-sequence lifetime of about 10×10^9 yr.

This page titled 43.E: Nuclear Physics (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44: PARTICLE PHYSICS AND COSMOLOGY

On a clear, moonless night, find the faintest smudge of light you can. If you are lucky enough to be looking toward a distant quasar, the photons now landing on your retina began their journey billions of years ago, when the universe was a fraction of its present age: they have been crossing empty, expanding space ever since. At the very same instant, trillions of neutrinos from the Sun, from exploding stars, and from the Big Bang itself are streaming silently through your body, almost none of them stopping. The light tells you the universe is vast and old and growing; the neutrinos tell you it is built from particles so ghostly we needed half a century to be sure they had any mass at all.

These two threads (*what the universe is made of* and *where it came from and where it is going*) are the subject of this final chapter, and they turn out to be the same story told at two scales. The same physics that governs a quark inside a proton governed the entire cosmos in its first microsecond. We will start small, cataloging the elementary particles and the forces that bind them into the Standard Model, the most precisely tested theory in the history of science. Then we will zoom outward to the expanding universe, to the afterglow of its hot birth, and to the two great mysteries (dark matter and dark energy) that tell us our beautiful Standard Model describes only about one part in twenty of everything there is. The strong force that held the nucleus together in Chapter 43 will reappear here as a residue of a deeper force between quarks.

Learning Objectives

- Classify particles as fermions or bosons by their spin, and as hadrons or leptons by whether the strong force acts on them.
- Describe antiparticles and the conservation laws (charge, baryon number, lepton number, strangeness) that govern particle reactions, and use them to decide whether a reaction can occur.
- Build hadrons from quarks and read off their charge, strangeness, and baryon number.
- Identify the four fundamental forces, their messenger particles, and how the Standard Model and the Higgs mechanism organize them.
- Explain how a collider detector identifies particles, measure momentum from track curvature, and use the invariant mass to reconstruct a short-lived particle.
- Use Hubble's law to relate a galaxy's recession speed to its distance and to estimate the age of the universe.
- Explain the evidence for the hot Big Bang (the cosmic microwave background), for dark matter (galactic rotation curves and lensing), and for dark energy (accelerating expansion), and outline the possible ultimate fates of the universe.

We will return to that distant quasar, and to the question of what the universe is made of and how old it is, in the capstone problem at the end of the chapter.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

44.1: Elementary Particles
44.2: Basic Forces and Messenger Particles
44.3: The Standard Model
44.4: Cosmology
44.S: Particle Physics and Cosmology (Summary)
44.E: Particle Physics and Cosmology (Problems)

This page titled 44: Particle Physics and Cosmology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.1: Elementary Particles



Figure 44.1. The Standard Model of elementary particles. The fermions (matter) fall into three generations: two rows of *quarks* (purple) and two rows of *leptons* (green). The force-carrying *gauge bosons* (orange) and the *Higgs* scalar boson (gold) complete the picture. Masses are quoted as rest energies; neutrino masses are tiny upper limits (for the electron neutrino, below $0.45\ \text{eV}$, set by the KATRIN experiment in 2025).

•Figure description

The Standard Model laid out as a grid of colored cells. The left three columns are the three generations of matter, each column heavier than the one to its left. The top two rows, in purple, are the six quarks: up, charm and top on one row, down, strange and bottom on the next, labeled QUARKS down the side. The bottom two rows, in green, are the six leptons: electron, muon and tau, then their three neutrinos, labeled LEPTONS. A fourth column in orange, labeled GAUGE BOSONS, holds the four gauge bosons: gluon, photon, Z and W. A fifth column, labeled SCALAR BOSON, holds a single gold cell of the same size at the top right: the Higgs. Each cell gives the particle's symbol, name and mass as a rest energy, ranging from under half an electron volt for the electron neutrino to 173 GeV for the top quark.

When physicists first began smashing particles together at high energy, they expected to find a few simple building blocks. Instead, the debris of those collisions revealed several *hundred* distinct types of particles. Faced with this "particle zoo," physicists did what biologists and chemists had done before them: they looked for simple physical criteria by which to classify. That classification, refined over decades, is the **Standard Model** of particle physics (Figure 44.1).

Three classification questions organize almost everything:

- Is the particle a **fermion** or a **boson**? (Set by its spin.)
- Is it a **hadron** or a **lepton**? (Set by whether the strong force acts on it.)
- Is it a **particle** or an **antiparticle**?

We take these in turn, and they will lead us naturally to the quarks and to the forces that bind everything together. Figure 44.2 previews where the answers lead for the elementary particles themselves.

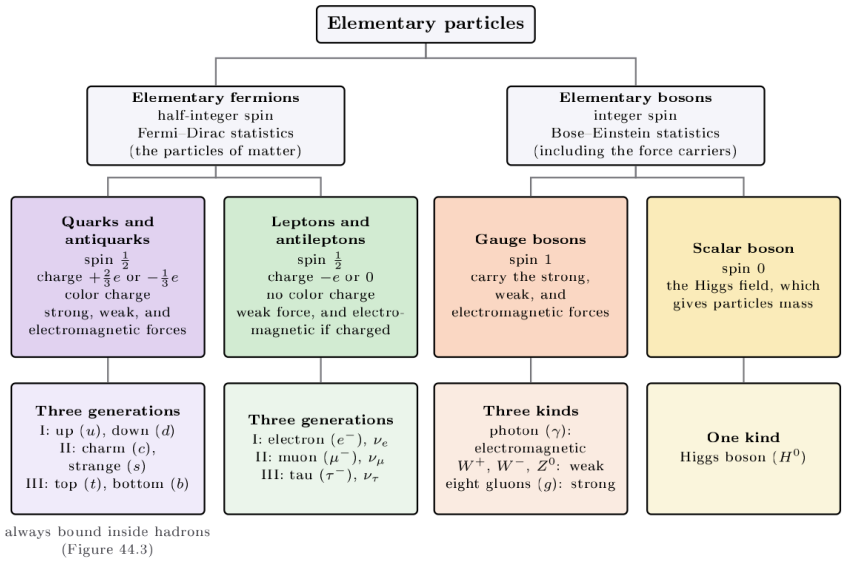


Figure 44.2. Classifying the elementary particles. Spin divides them into fermions and bosons. The fermions are the quarks and leptons, each in three generations; every one has an antiparticle. The bosons are the gauge bosons, which carry the strong, weak, and electromagnetic forces, and the Higgs boson. Compare the grid of Figure 44.1.

•Figure description

A tree diagram classifying the elementary particles. At the top, Elementary particles splits into two branches. The left branch, Elementary fermions, has half-integer spin, obeys Fermi-Dirac statistics and makes up matter. The right branch, Elementary bosons, has integer spin and obeys Bose-Einstein statistics. The fermions split into two purple and green boxes. Quarks and antiquarks: spin one half, fractional charge of plus two thirds or minus one third of e, carry color charge, and feel the strong, weak and electromagnetic forces. Leptons and antileptons: spin one half, charge minus e or zero, no color charge, and feel the weak force, plus the electromagnetic force if charged. The bosons split into two orange and gold boxes. Gauge bosons: spin 1, the force carriers. Scalar boson: spin 0, the Higgs field that gives particles mass. The bottom row lists the members. Quarks come in three generations: up and down, charm and strange, top and bottom; a note says they are always bound inside hadrons. Leptons come in three generations: electron and electron neutrino, muon and muon neutrino, tau and tau neutrino. Gauge bosons come in three kinds: the photon for the electromagnetic force, the W plus, W minus and Z zero for the weak force, and eight gluons for the strong force. There is one scalar boson, the Higgs boson, H zero.

Fermion vs Boson

Every particle carries an intrinsic angular momentum called **spin**, introduced in our study of atomic structure (Chapter 41). Spin is quantized: the component of spin along any axis is

Key Equation: Quantized spin component

$$S_z = m_s\,\hbar, \qquad m_s = -s,\ -s+1,\ \ldots,\ +s, \tag{44.1}$$

where $\hbar = h/2\pi$, s is the *spin quantum number* of the particle, and m_s is the *spin magnetic quantum number*. An electron, for example, has $s = \frac{1}{2}$, so $m_s = \pm\frac{1}{2}$ and $S_z = \pm\frac{1}{2}\hbar$: the familiar "spin up" and "spin down" states.

The single most important fact about a particle's spin is whether it is a half-odd-integer or a whole number, because that determines its statistics:

- **Fermions** have half-odd-integer spin ($s = \frac{1}{2}, \frac{3}{2}, \ldots$) and obey *Fermi–Dirac statistics*. Examples: electrons, protons, neutrons, quarks ($s = \frac{1}{2}$). Fermions are further divided into leptons and quarks.
- **Bosons** have integer spin ($s = 0, 1, 2, \ldots$) and obey *Bose–Einstein statistics*. Examples: photons and gluons ($s = 1$), the Higgs ($s = 0$). The elementary bosons act as force carriers.

Why insist on this division? Because the two families behave in opposite ways when crowded together. Fermions obey the **Pauli exclusion principle**: no two identical fermions can occupy the same quantum state. This is exactly the rule that forced electrons

to stack into successive shells in atoms (Chapter 41) and that gives matter its bulk and rigidity. Bosons obey no such restriction: any number of them can pile into a single quantum state.

That bosonic freedom has a spectacular consequence. Cool a dilute gas of bosonic atoms to within a whisker of absolute zero and a large fraction of them suddenly crowd into the *single lowest* quantum state, forming a **Bose–Einstein condensate** (BEC): sometimes called the fifth state of matter. In a BEC, microscopic quantum behavior such as wavefunction interference becomes visible on a macroscopic scale, as if the whole cloud of atoms had merged into one giant quantum wave.

Checkpoint 44.1

A helium-4 atom contains 2 protons, 2 neutrons, and 2 electrons: six spin-$\frac{1}{2}$ fermions. Is the atom as a whole a fermion or a boson? (Hint: an even number of half-integer spins combines to an integer total.) Helium-4 famously *can* be cooled into a superfluid; does your answer help explain why?

Answer

Particles are also sorted by *which fundamental forces act on them*. Gravity acts on everything with energy, but is so feeble at the particle scale that we ignore it. The electromagnetic force acts on all electrically charged particles. The weak force acts on essentially all particles. And the strong force acts *only on quarks*: a fact that gives us our next, and most useful, classification.

Hadron vs Lepton

The cleanest line we can draw through the particle zoo is this: does the strong force act on the particle or not?

- **Hadrons** are particles that *do* feel the strong force: protons, neutrons, pions, kaons. They feel it because they are built from quarks.
- **Leptons** are particles that do *not* feel the strong force: electrons, muons, taus, and the neutrinos. Leptons appear to be truly fundamental, with no internal structure.

Quarks are the only fundamental particles acted on by all four forces, including the strong force, and the only known particles carrying electric charge in non-integer multiples of the elementary charge e (they carry $\pm\frac{1}{3}e$ or $\pm\frac{2}{3}e$).

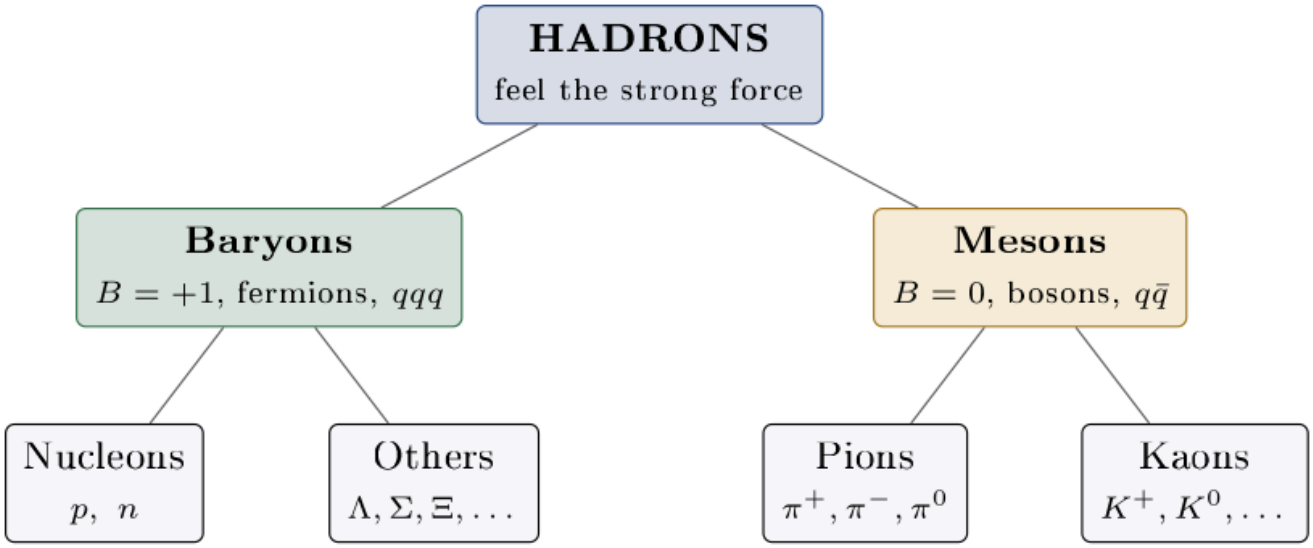


Figure 44.3. The hadron family. All hadrons feel the strong force because all are built from quarks. Baryons (three quarks, baryon number $B = +1$) are fermions; mesons (a quark–antiquark pair, $B = 0$) are bosons.

•Figure description

A branching tree diagram of the hadron family. A blue box at the top reads HADRONS, feel the strong force. It splits into two boxes: Baryons on the left, described as baryon number plus one, fermions, three quarks; and Mesons on the right, baryon number zero, bosons, a quark and an antiquark. Baryons then splits into Nucleons, the proton and neutron, and Others such as lambda, sigma and xi. Mesons splits into Pions and Kaons.

Hadrons split further according to whether they are bosons or fermions (Figure 44.3):

- **Mesons** are hadrons that are bosons. Each is built from *one quark and one antiquark*. Examples: the pions π^+, π^-, π^0 and the kaons K^+, K^0.
- **Baryons** are hadrons that are fermions. Each is built from an *odd* number of valence quarks, almost always three. Examples: the proton (uud) and neutron (udd).

Checkpoint 44.2

A newly observed particle feels the strong force and has spin $s = \frac{3}{2}$. Into which branch of Figure 44.3 must it fall, and how many valence quarks should it contain?

Answer

Antiparticles

For nearly every particle there exists an **antiparticle** with the *same mass and spin* but the *opposite sign* of every charge-like quantum number: electric charge, color charge, strangeness, baryon number, and lepton number. We denote an antiparticle with a bar over its symbol: the antiproton is $\bar{p}$, the anti–up-quark is $\bar{u}$. (A few neutral particles, such as the photon, are their own antiparticles.) The positron e^+, the electron's antiparticle, was predicted by Dirac's relativistic theory and discovered in cosmic rays in 1932: the first evidence that antimatter is real.

When a particle meets its antiparticle, they can **annihilate**: both vanish and their combined energy reappears in other forms, often as photons or as new particle–antiparticle pairs. Because mass is a form of energy (Chapter 35), the minimum energy released when a particle of rest energy mc^2 annihilates with its antiparticle is

Key Equation: Particle–antiparticle annihilation energy

$$E_{\text{annih}} = 2mc^2, \tag{44.2}$$

shared among the products. The reverse process, *pair production*, turns energy back into a particle–antiparticle pair, and is the everyday business of high-energy colliders.

Every reaction among particles must respect a set of **conservation laws**. In addition to energy and momentum, the following quantities are conserved (the last, strangeness, only in the strong and electromagnetic interactions, not the weak):

- electric charge q;
- baryon number B ($+1$ for each baryon, -1 for each antibaryon, 0 for mesons and leptons);
- lepton number, counted *separately* for each family (L_e, L_μ, L_τ);
- strangeness S (strong and EM interactions only).

We can already use energy and momentum conservation to analyze a decay completely.

Example 44.1: Energy sharing in pion decay

A stationary positive pion decays according to $\pi^+ \to \mu^+ + \nu$. Find the kinetic energy of the antimuon μ^+ and of the neutrino ν. Take the rest energies to be $m_\pi c^2 = 139.6\ \text{MeV}$ and $m_\mu c^2 = 105.7\ \text{MeV}$, and treat the neutrino as massless.

Solution

The pion is at rest, so the total momentum is zero: the muon and neutrino fly apart with equal and opposite momenta of magnitude p. Energy conservation gives

$$m_\pi c^2 = E_\mu + E_\nu, \tag{44.1.1}$$

with $E_\mu = \sqrt{(pc)^2 + (m_\mu c^2)^2}$ for the massive muon and $E_\nu = pc$ for the massless neutrino (Chapter 35). Substituting and isolating the square root,

$$\sqrt{(pc)^2 + (m_\mu c^2)^2} = m_\pi c^2 - pc. \tag{44.1.2}$$

Squaring, the $(pc)^2$ terms cancel and we can solve for pc:

$$pc = \frac{(m_\pi c^2)^2 - (m_\mu c^2)^2}{2\, m_\pi c^2} = \frac{(139.6)^2 - (105.7)^2}{2(139.6)}\ \text{MeV} = 29.8\ \text{MeV}. \tag{44.1.3}$$

The neutrino is massless, so all of its energy is kinetic:

$$K_\nu = E_\nu = pc = \boxed{29.8\ \text{MeV}}. \tag{44.1.4}$$

The muon's total energy is $E_\mu = m_\pi c^2 - E_\nu = 139.6 - 29.8 = 109.8\ \text{MeV}$, so its kinetic energy is

$$K_\mu = E_\mu - m_\mu c^2 = 109.8 - 105.7 = \boxed{4.1\ \text{MeV}}. \tag{44.1.5}$$

As a check, the two kinetic energies are wildly unequal even though the momenta are equal in magnitude, just as expected, because the light, ultrarelativistic neutrino carries far more kinetic energy per unit momentum than the heavy, slow muon. Their kinetic energies sum to $4.1 + 29.8 = 33.9\ \text{MeV}$, which equals $m_\pi c^2 - m_\mu c^2$: the rest energy "freed up" by the decay, as it must.

Quarks

The hadrons are built from six **quarks**, in three generations of increasing mass: up and down; charm and strange; top and bottom. Every quark is a spin-$\frac{1}{2}$ fermion, so every antiquark is too. Each quark carries a fractional electric charge, a baryon number of $+\frac{1}{3}$, and (for the strange quark) a unit of strangeness. The quantum numbers of an antiquark are the negatives of those of the corresponding quark. Table 44.1 collects the values.

Table 44.1. The six quarks and their quantum numbers. Masses are approximate, and match Figure 44.1. Each antiquark ($\bar{u}, \bar{d}, \ldots$) has the opposite signs of q, S, and B. By convention the strange quark carries strangeness $S = -1$.

Quark	Symbol	Mass (MeV/c^2)	Charge q	Strangeness S	Baryon B
Up	u	≈ 2.2	$+\frac{2}{3}$	0	$+\frac{1}{3}$
Down	d	≈ 4.7	$-\frac{1}{3}$	0	$+\frac{1}{3}$
Charm	c	≈ 1270	$+\frac{2}{3}$	0	$+\frac{1}{3}$
Strange	s	≈ 93	$-\frac{1}{3}$	-1	$+\frac{1}{3}$
Top	t	$\approx 173\ 000$	$+\frac{2}{3}$	0	$+\frac{1}{3}$
Bottom	b	≈ 4180	$-\frac{1}{3}$	0	$+\frac{1}{3}$

To build a hadron, combine quarks so that their charges, strangenesses, and baryon numbers add to those of the hadron. A baryon is three quarks ($B = 3 \times \frac{1}{3} = +1$); a meson is a quark plus an antiquark ($B = \frac{1}{3} - \frac{1}{3} = 0$). The proton is uud (charge $\frac{2}{3} + \frac{2}{3} - \frac{1}{3} = +1$); the neutron is udd (charge $\frac{2}{3} - \frac{1}{3} - \frac{1}{3} = 0$); the positive pion is $u\bar{d}$.

Example 44.2: Quark content of the Ξ^-

The Ξ^- ("xi-minus") is a baryon with spin $s = \frac{1}{2}$, charge $q = -1$, and strangeness $S = -2$. What combination of quarks makes it up?

Solution

A baryon contains three quarks, and we need them to deliver $q = -1$, $S = -2$, and $B = +1$. The two units of (negative) strangeness require *two strange quarks*, since only the s quark carries strangeness, at $S = -1$ each:

$$S = (-1) + (-1) + S_{\text{third}} = -2 \;\Rightarrow\; S_{\text{third}} = 0, \tag{44.1.6}$$

so the third quark is not strange. Now match the charge. The two strange quarks contribute $2 \times (-\frac{1}{3}) = -\frac{2}{3}$, so the third quark must supply

$$q_{\text{third}} = -1 - \left(-\tfrac{2}{3}\right) = -\tfrac{1}{3}, \tag{44.1.7}$$

which is a down quark. The composition is therefore

$$\boxed{\Xi^- = d\,s\,s}. \tag{44.1.8}$$

Check: charge $-\frac{1}{3} - \frac{1}{3} - \frac{1}{3} = -1$ ✓; strangeness $0 - 1 - 1 = -2$ ✓; baryon number $3 \times \frac{1}{3} = +1$ ✓. Three half-integer spins can combine to a half-integer total, consistent with $s = \frac{1}{2}$.

All of these quantum numbers must be conserved in a reaction, but with a crucial caveat: *strangeness is conserved by the strong and electromagnetic interactions, yet can change by one unit in a weak interaction*. This single rule explains why "strange" particles are produced copiously (in pairs, by the strong force, conserving S) yet decay slowly (one at a time, by the weak force, violating S). The next example puts the full set of conservation laws to work.

Example 44.3: Which reactions can occur?

Use conservation laws to decide whether each reaction is allowed. (Treat each as a strong or electromagnetic process unless a particle simply decays.)

a. $p + n \to p + p + n + \bar{p}$
b. $p + n \to p + p + \bar{p}$
c. $\mu^- \to e^- + \bar{\nu}_e + \nu_\mu$
d. $\pi^+ \to \mu^+ + \nu_\mu + \nu_e$
e. $\pi^0 + n \to K^+ + \Sigma^-$

f. $\pi^- + p \to \pi^- + \Sigma^+$

Solution

We check charge q, baryon number B, the lepton numbers L_e, L_μ, and (for hadronic reactions) strangeness S.

(a) Charge: $1 + 0 = 1 + 1 + 0 - 1$, i.e. $1 = 1$ ✓. Baryon number: $1 + 1 = 1 + 1 + 1 - 1$, i.e. $2 = 2$ ✓. No leptons or strangeness involved. **Allowed**: it creates a proton–antiproton pair, possible above threshold energy.

(b) Charge: $1 = 1 + 1 - 1 = 1$ ✓. Baryon number: $1 + 1 = 2$ on the left but $1 + 1 - 1 = 1$ on the right. Since $2 \neq 1$, B is not conserved. **Forbidden**.

(c) This is the standard muon decay. Charge: $-1 = -1$ ✓. Electron number: left 0; right $(+1)_{e^-} + (-1)_{\bar{\nu}_e} + 0 = 0$ ✓. Muon number: left $(+1)_{\mu^-}$; right $0 + 0 + (+1)_{\nu_\mu} = +1$ ✓. **Allowed**.

(d) Charge: $+1 = +1$ ✓. Muon number: left 0; right $(-1)_{\mu^+} + (+1)_{\nu_\mu} + 0 = 0$ ✓. Electron number: left 0; right $0 + 0 + (+1)_{\nu_e} = +1 \neq 0$. Electron lepton number is violated. **Forbidden**. (The real decay is $\pi^+ \to \mu^+ + \nu_\mu$, with no ν_e.)

(e) A hadronic (strong) reaction, so strangeness must be conserved. Charge: $0 + 0 = (+1) + (-1) = 0$ ✓. Baryon number: $0 + 1 = 0 + 1$ ✓. Strangeness: $K^+ = u\bar{s}$ has $S = +1$ and $\Sigma^- = dds$ has $S = -1$, so the right side has $S = +1 - 1 = 0$, equal to the left. ✓ **Allowed**.

(f) Charge: $-1 + 1 = -1 + 1 = 0$ ✓. Baryon number: $0 + 1 = 0 + 1$ ✓. Strangeness: the left side has $S = 0$; on the right $\Sigma^+ = uus$ has $S = -1$, so S would change by one unit. A strong production reaction cannot change strangeness. **Forbidden** (as a strong reaction).

Answers to Checkpoints

Checkpoint 44.1. A boson. Six spin-$\frac{1}{2}$ particles combine to an integer total spin (for helium-4 in its ground state, zero). Bosons are not bound by the exclusion principle, so any number of them can crowd into the same lowest-energy quantum state. When helium-4 is cooled enough, a large fraction of its atoms do exactly that, moving together as one quantum state that flows without friction: a superfluid. (Helium-3, with an odd number of fermions, is a fermion and becomes superfluid only at far lower temperatures, by first pairing up.)

Checkpoint 44.2. Feeling the strong force makes it a hadron, and a half-integer spin makes it a fermion. So it must be a *baryon*, and baryons contain three valence quarks. (Mesons, the other hadron branch, have integer spin and a quark–antiquark pair.)

This page titled 44.1: Elementary Particles is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.2: Basic Forces and Messenger Particles

Everything in the previous sections is held together by just *four* fundamental forces: gravity, the electromagnetic force, the weak force, and the strong force. Modern physics describes each (except, so far, gravity) as the exchange of a **messenger particle**, or force carrier: a boson passed back and forth between the interacting particles.

The Electromagnetic Force

Two electrons near each other repel through Coulomb's law,

$$F_C = \frac{k\,q_1 q_2}{r^2}, \tag{44.3}$$

but at the quantum level the interaction is described by *quantum electrodynamics* (QED). In that picture each electron senses the other by exchanging photons, which exist only fleetingly and are therefore called **virtual photons**. Because they carry the interaction between particles, such exchanged bosons are the messenger particles, or force carriers, of the force (Figure 44.4).

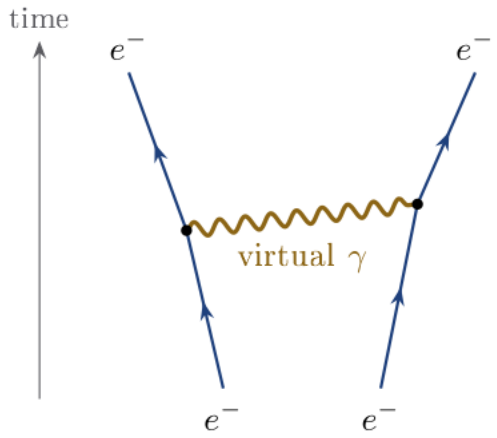


Figure 44.4. Electric repulsion as quantum electrodynamics describes it. Two electrons exchange a virtual photon, and each is deflected; time runs upward. Compare the beta-decay diagram of Figure 44.5, where the exchanged boson is a W^-.

•Figure description

A Feynman diagram of two electrons repelling, with time running upward. Two electron lines come in from the bottom, labeled e minus. Each reaches a vertex, the two vertices joined by a gold wavy line labeled virtual gamma. After the exchange the two electron lines leave toward the top, bent away from each other, still labeled e minus.

The Weak Force

The weak force, responsible for beta decay (Chapter 43), is described by a theory so closely analogous to QED that physicists eventually realized the electromagnetic and weak forces are two aspects of a single **electroweak** force. Its messengers are not the massless photon but three *massive* bosons: the W^+, the W^-, and the Z^0. Their large mass is exactly why the weak force has such a short range. Figure 44.5 shows the weak force at work in the most familiar case, the beta decay of a neutron.

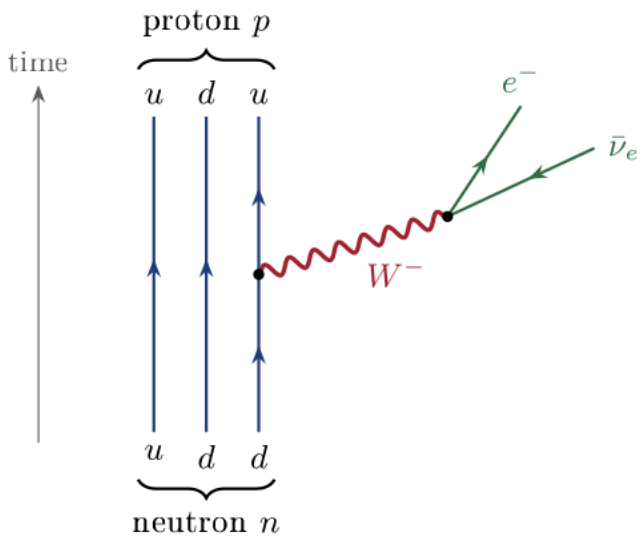


Figure 44.5. Beta decay, $n \to p + e^- + \bar{\nu}_e$, seen at the quark level. One down quark in the neutron emits a W^- boson and becomes an up quark, turning udd into uud; the short-lived W^- then produces the electron and its antineutrino. Time runs upward; the arrow on the antineutrino line points backward, the standard way of drawing an antiparticle.

•Figure description

A Feynman diagram of beta-minus decay, with time running upward. At the bottom, three quark lines labeled u, d and d enter together, bracketed as a neutron. Two of the lines, u and d, run straight up unchanged. The third, a d quark, reaches a vertex where it turns into a u quark and emits a wavy line labeled W minus, heading off to the right. At the top the three lines u, d and u leave together, bracketed as a proton. The W minus travels a short way and ends at a second vertex, where it becomes an outgoing electron and an outgoing electron antineutrino, both heading up and to the right.

The Strong Force

The strong force binds quarks into hadrons; its carriers are **gluons**, which are massless like photons. Quarks carry a kind of charge called **color charge**, which comes in three varieties, conventionally labeled *red*, *green*, and *blue*; each antiquark carries one of the corresponding anticolors. Quarks combine only in *color-neutral* ("white") arrangements:

- three quarks, one of each color: $R + G + B = \text{white}$ (a baryon, a fermion);
- a quark and an antiquark of matching color/anticolor: $R + \bar{R} = \text{white}$ (a meson, a boson).

The theory of the color force is *quantum chromodynamics* (QCD), built in deliberate analogy to QED. The same color force, acting between the quarks of neighboring nucleons, leaks out as the residual "nuclear" strong force that binds protons and neutrons together, the very force we invoked in Chapter 43.

Table 44.2 compares the three forces of the Standard Model and their carriers.

Table 44.2. The three Standard-Model forces and their messenger particles. The strong force is confined to within a hadron; the weak force is even shorter-ranged because its carriers are massive. Gravity, with no confirmed quantum carrier, is omitted.

Property	Electromagnetic	Strong	Weak
Force carrier	photon (γ)	gluon (g)	W^+, W^-, Z^0
Carrier mass	0	0	large
Charge of carrier	none	color	weak ($W^\pm$ also electric)
Couples to	charged particles	quarks only	all quarks & leptons
Range	infinite ($\propto 1/r^2$)	$\approx 10^{-15}$ m	$\approx 10^{-18}$ m

The Higgs Field

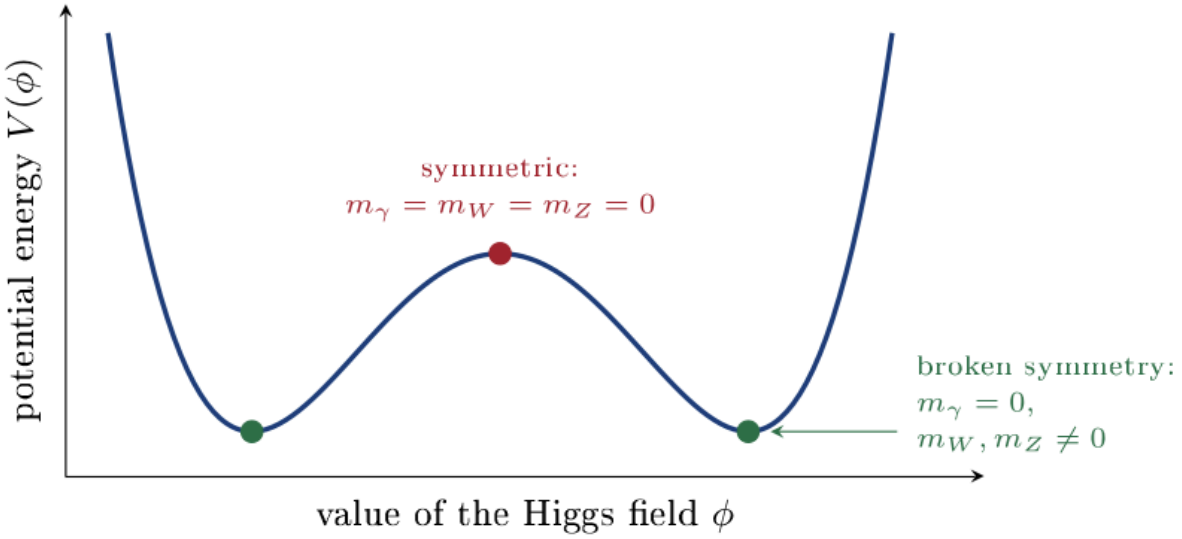


Figure 44.6. A cross-section of the Higgs potential (the famous “Mexican hat”). The symmetric state at the central peak (red) would leave all electroweak carriers massless; the field instead settles into one of the lower-energy minima (green), breaking the symmetry and endowing the W and Z with mass while the photon stays massless.

•Figure description

A cross-section of the Higgs potential, the famous Mexican-hat shape. The curve of potential energy against the value of the Higgs field is a W: it has a local bump in the center and two equal minima either side of it. A red dot sits on top of the central bump, labeled symmetric, where the photon, W and Z would all be massless. Green dots sit in the two minima, labeled broken symmetry, where the photon stays massless but the W and Z acquire mass. An arrow shows the field rolling off the unstable central peak down into one of the minima.

A deep puzzle remains: why are the photon and gluon massless while the W and Z are enormously heavy? The resolution is the **Higgs field**, a field proposed to permeate all of space: a property of the vacuum itself. In the Brout–Englert–Higgs mechanism, this field *spontaneously breaks* the symmetry of the electroweak force: without it, all the electroweak carriers would be massless and indistinguishable, but the field selects a particular vacuum state, giving the W and Z their masses while leaving the photon massless (Figure 44.6). The same field gives the quarks and the charged leptons their masses. The quantum of this field is the **Higgs boson**, discovered in 2012.

Checkpoint 44.3

The weak and electromagnetic forces have very different ranges in everyday experience, yet they are “unified.” Using Table 44.2, explain how the *mass* of the force carrier (rather than any difference in the underlying theory) accounts for the weak force’s tiny range.

Answer

Answers to Checkpoints

Checkpoint 44.3. The photon, which carries the electromagnetic force, is massless, while the W and Z, which carry the weak force, have masses of about 80 and $91\ \mathrm{GeV}/c^2$. A force carrier exchanged between particles is a virtual particle that must “borrow” energy $\Delta E \approx Mc^2$, and the uncertainty principle lets that loan last only $\Delta t \sim \hbar/Mc^2$. In that time it travels at most about $c\ \Delta t \sim \hbar/Mc$, roughly $10^{-18}\ \mathrm{m}$ for the W. A massless photon needs no such loan, so the electromagnetic force has infinite range. The theory is the same; the carriers’ masses, given by the Higgs field, make the difference.

This page titled 44.2: Basic Forces and Messenger Particles is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.3: The Standard Model

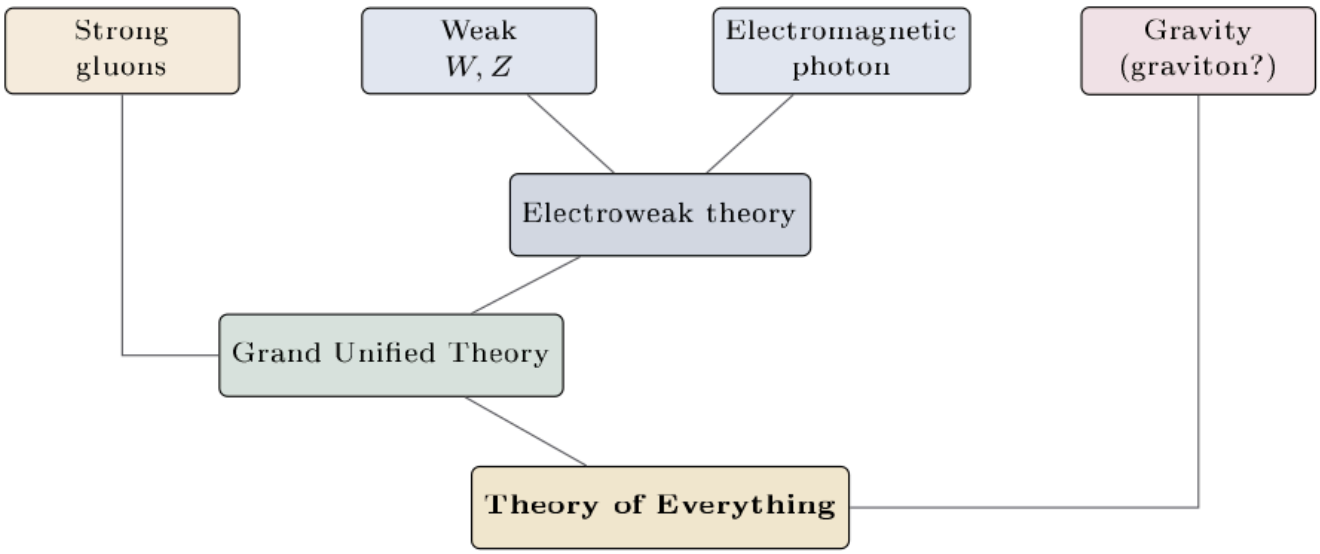


Figure 44.7. The ladder of unification. The electromagnetic and weak forces are already joined into the electroweak theory. Grand Unified Theories aim to add the strong force; a Theory of Everything would finally bring in gravity. Only the electroweak step is firmly established.

•Figure description

A ladder of unification, drawn as boxes joined by lines and read downward. Across the top are the four fundamental forces with their carriers: Strong with gluons, Weak with W and Z, Electromagnetic with the photon, and Gravity with a question-marked graviton. Weak and Electromagnetic join into a single box, Electroweak theory. That box and Strong join into Grand Unified Theory. Finally Grand Unified Theory and Gravity join into Theory of Everything at the bottom. Only the first step, electroweak, is established.

We can now assemble the whole picture. Leptons, quarks, and the force-carrying bosons are **fundamental particles**: as far as every experiment can tell, they have no internal structure. From them, the composite particles (hadrons, then nuclei, then atoms, then molecules) are built up in layers. The forces, meanwhile, have been progressively *unified* (Figure 44.7).

The unification of the electromagnetic and weak forces into the electroweak force is established and experimentally confirmed. Beyond it, physicists pursue **Grand Unified Theories** (GUTs) that would fold in the strong force as well. And conspicuously absent from the Standard Model is *gravity*: it is described instead by Einstein's general relativity (Chapter 35), and no one has yet reconciled that geometric theory with the quantum framework of the other three forces. A successful merger would be a **Theory of Everything**; candidate frameworks include string theory and loop quantum gravity, but none is yet established.

Checkpoint 44.4

(a) The Δ^{++} is a baryon with charge $+2e$ and no strangeness. What must its quark content be? (b) Which of these are fundamental particles, and which are composite: electron, proton, photon, neutron, up quark, pion?

Answer

Detecting Particles

How do we know that any of these particles exist? Almost none of them can be seen directly. A Higgs boson lives for about 2×10^{-22} s and a Z boson for even less; in that time, even at nearly the speed of light, they travel far less than the width of an atom before decaying. A quark is never seen alone at all. What a detector actually records are the handful of long-lived particles the short-lived ones decay into: electrons, photons, muons, neutrinos, and hadrons such as pions, kaons, protons, and neutrons. Discovering a new particle means working backward from those debris to the particle that made them.

The particles are produced in collisions. Two beams of protons, each accelerated to several TeV in a synchrotron (Section 22.3), are steered into each other inside a detector. In the collision the protons' kinetic energy becomes available to create new particles of mass up to E/c^2, by equation (35.10). The Large Hadron Collider near Geneva brings its beams together at four points, and at two of them stand the general-purpose detectors ATLAS and CMS, each the size of a building.

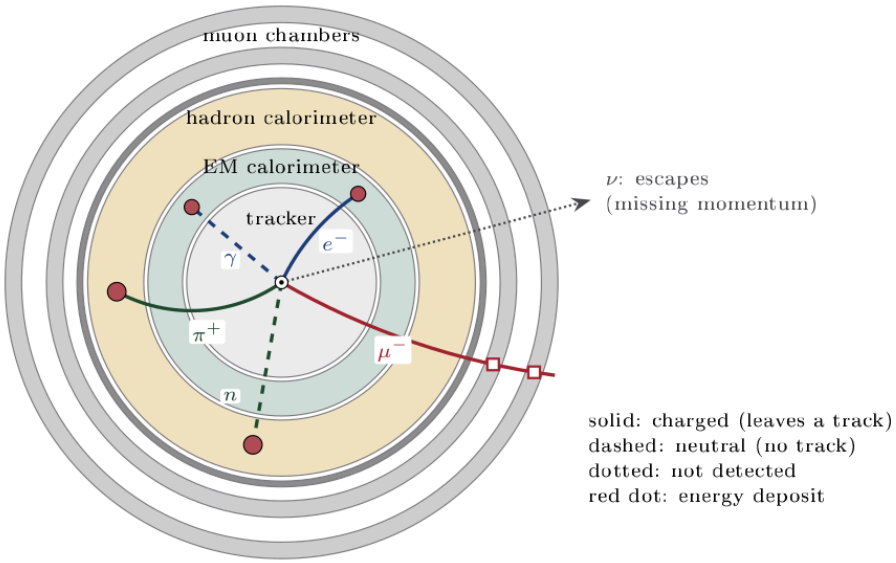


Figure 44.8. An end-on view of a general-purpose collider detector (schematic, not to scale), looking along the beams, which collide at the center. Each kind of particle leaves its own signature. Charged particles curve in the magnetic field of the tracker; electrons and photons stop in the electromagnetic (EM) calorimeter, and hadrons in the hadron calorimeter; muons pass through everything to the muon chambers; neutrinos escape unseen and are revealed only by the momentum they carry away.

•Figure description

An end-on view of a collider detector, looking along the beam, which runs into the page at the center, where the collision happens. Concentric rings surround the center. From the inside out they are: the tracker, a pale gray disk; the electromagnetic calorimeter, a green ring; the hadron calorimeter, a gold ring; a thin dark ring for the solenoid magnet; and two gray rings of muon chambers. Six particles leave the center. An electron, drawn as a solid blue curved track, bends in the magnetic field and stops in the electromagnetic calorimeter, where a red blob marks its deposited energy. A photon, drawn dashed because it leaves no track, travels straight to a deposit in the electromagnetic calorimeter. A charged pion, a solid curved track, stops in the hadron calorimeter with a deposit there. A neutron, dashed and straight, also stops in the hadron calorimeter. A muon, a solid and only slightly curved track, passes through every layer and leaves hits in both muon chambers. A neutrino, drawn as a dotted arrow, passes out through everything, labeled missing momentum.

A modern detector is built like an onion, in layers wrapped around the collision point, each designed to catch a different kind of particle (Figure 44.8).

- The innermost layer, the **tracker**, is made of thin silicon sensors that record the points where a charged particle passes, with a precision of about $10\ \mu\text{m}$. It sits inside a large solenoid that fills it with a strong magnetic field ($3.8\ \text{T}$ in CMS), so the tracks of charged particles curve. Neutral particles pass through the tracker without leaving a trace.
- A **calorimeter** measures energy by stopping a particle completely. The *electromagnetic calorimeter* absorbs electrons and photons in a shower of secondary electrons, positrons, and photons, and measures the total energy deposited. The thicker *hadron calorimeter* behind it absorbs protons, neutrons, pions, and the other hadrons. A quark or gluon knocked out in the collision appears as a **jet**: a narrow spray of hadrons, created as the strong force refuses to let the quark escape on its own.
- Muons, which do not feel the strong force and are too massive to be stopped easily by the electromagnetic interactions that halt electrons, punch through everything. The outermost layers are **muon chambers**, and anything that reaches them is almost certainly a muon.
- Neutrinos interact so weakly that they escape without leaving any signal at all.

Putting the layers together, each particle has a signature: an electron leaves a track and stops in the electromagnetic calorimeter; a photon stops there too but leaves no track; a charged hadron leaves a track and stops in the hadron calorimeter; a muon leaves a track all the way out.

Measuring momentum. The curvature of a track measures the particle's momentum. A charged particle moving perpendicular

to a magnetic field travels in a circle of radius $r = p/|q|B$ (equation (22.2)), and this remains exact for relativistic particles when $p = \gamma m v$. A slow particle curls up tightly, while a very energetic one leaves an almost straight track; the direction of the bend reveals the sign of the charge. For a particle of charge $\pm e$ it is convenient to write the result in the units of particle physics,

$$pc = ecBr = (0.2998\,\text{GeV})\left(\frac{B}{1\,\text{T}}\right)\left(\frac{r}{1\,\text{m}}\right), \tag{44.4}$$

since $ecBr$ for $B = 1\ \text{T}$ and $r = 1\ \text{m}$ is $(1.602 \times 10^{-19}\ \text{C})(2.998 \times 10^{8}\ \text{m/s})(1\ \text{T})(1\ \text{m}) = 4.80 \times 10^{-11}\ \text{J} = 0.2998\ \text{GeV}$.

Missing momentum. The colliding protons move along the beam line, so before the collision there is no momentum at all *across* the beam. By conservation of momentum, the momenta of everything produced must also add to zero in the plane perpendicular to the beam. If the visible particles do not balance, something invisible carried the difference away. This **missing transverse momentum** is how a detector "sees" a neutrino, and it is also how physicists search for undiscovered neutral particles such as candidates for dark matter (Section 44.4).

Reconstructing a decay. To find a particle that decays before it can be detected, physicists add up the energies and momenta of its decay products. By equation (35.12), a particle's energy and momentum satisfy $E^2 = (pc)^2 + (mc^2)^2$, so its mass follows from the *total* energy and momentum of the products, whatever the speed of the decaying particle:

Key Equation: Invariant mass of decay products

$$(Mc^2)^2 = \left(\sum E\right)^2 - \left|\sum \mathbf{p}\,c\right|^2. \tag{44.5}$$

This quantity, the **invariant mass**, is the same in every reference frame. Each pair of photons (or electrons, or muons) recorded in a detector gives one value of it. Pairs that merely happen to occur together give values scattered smoothly over a wide range, but pairs that come from the decay of one particle all give the same value: its mass. A new particle shows up as a narrow bump standing above the smooth background in a histogram of invariant masses. In 2012 the ATLAS and CMS experiments each found such a bump at $125\ \text{GeV}$, in pairs of photons and in sets of four leptons, after sifting through about a thousand trillion proton–proton collisions: the Higgs boson.

Example 44.4: Reconstructing a Higgs boson

A Higgs boson decays to two photons. The detector records one photon with energy $95.0\ \text{GeV}$ and the other with $55.0\ \text{GeV}$, with an angle of 120.0° between their directions. (a) Show that for two photons, equation (44.5) becomes $(Mc^2)^2 = 2E_1E_2(1 - \cos\theta)$, and find the invariant mass. (b) How fast was the Higgs boson moving?

Solution

(a) A photon has no mass, so $pc = E$ for each. The total energy is $E_1 + E_2$, and by the law of cosines the total momentum satisfies $|\sum \mathbf{p}\,c|^2 = E_1^2 + E_2^2 + 2E_1E_2\cos\theta$. Then

$$(Mc^2)^2 = (E_1 + E_2)^2 - \left(E_1^2 + E_2^2 + 2E_1E_2\cos\theta\right) = 2E_1E_2(1 - \cos\theta). \qquad \checkmark \tag{44.3.1}$$

With $\cos 120.0^\circ = -0.500$,

$$Mc^2 = \sqrt{2(95.0)(55.0)(1.500)}\,\text{GeV} = \sqrt{15{,}675}\,\text{GeV} = \boxed{125\,\text{GeV}}, \tag{44.3.2}$$

the mass of the Higgs boson.

(b) The Higgs boson's energy was $E = 95.0 + 55.0 = 150.0\ \text{GeV}$, and its momentum

$$pc = \sqrt{(95.0)^2 + (55.0)^2 + 2(95.0)(55.0)\cos 120.0^\circ}\,\text{GeV} = \sqrt{6825}\,\text{GeV} = 82.6\,\text{GeV}. \tag{44.3.3}$$

For any particle, $pc/E = \gamma m v c/(\gamma m c^2) = v/c$, so

$$v = \frac{pc}{E}\,c = \frac{82.6}{150.0}\,c = \boxed{0.551c}. \tag{44.3.4}$$

The Higgs was moving, so its photons did not fly apart back to back. The invariant mass does not care: it gives $125\ \mathrm{GeV}$ whatever the Higgs boson's speed.

Checkpoint 44.5

A collision produces a W^- boson, which decays to an electron and an antineutrino. Describe what the detector records. How can physicists tell that a neutrino was there at all?

Answer

Answers to Checkpoints

Checkpoint 44.4. (a) A baryon has three quarks, and with no strangeness they must come from u ($+\frac{2}{3}$) and d ($-\frac{1}{3}$). The only way three of these can total $+2$ is $\frac{2}{3}+\frac{2}{3}+\frac{2}{3}$, so the Δ^{++} is uuu. (Three identical quarks in the same state seemed to break the Pauli exclusion principle, a puzzle that helped lead physicists to color charge: the three quarks differ in color.) (b) Fundamental: the electron (a lepton), the photon (a force carrier), and the up quark. Composite: the proton and neutron (three quarks each) and the pion (a quark–antiquark pair).

Checkpoint 44.5. The detector records only the electron: a curved track in the tracker that ends in a deposit in the electromagnetic calorimeter. The antineutrino leaves no trace in any layer. But before the collision there was no momentum across the beam, so the momenta of everything produced must add to zero in that plane. The electron's transverse momentum is left unbalanced, and the *missing transverse momentum*, roughly equal and opposite to the electron's, reveals the antineutrino that carried it away.

This page titled 44.3: The Standard Model is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.4: Cosmology

Why dwell on whether the four forces unify? Because in the early universe they *did*. The four interactions were not always distinct. According to the Big Bang model, the universe began in an unimaginably hot, dense state. If the grand unified theories of the previous section are right, at the very earliest moments there was a single unified force, and as the universe expanded and cooled, the forces "froze out" one by one into the four we see today. **Cosmology** (the study of the universe as a whole, its origin, evolution, and fate) is therefore inseparable from particle physics. The hot early universe was the ultimate particle accelerator.



Figure 44.9. A timeline of the universe (not to scale). After the Big Bang and a brief inflationary burst, the universe expanded and cooled; at about 380,000 years it became transparent, releasing the cosmic microwave background. Stars and then galaxies formed within the first few hundred million years (the James Webb Space Telescope has seen galaxies from less than 300 million years after the Big Bang), and for roughly the last six billion years the expansion has been accelerating.

•Figure description

A horizontal timeline of the universe, not to scale, running left to right with six marked milestones. At the far left, t equals zero, the Big Bang and inflation. Then the cosmic microwave background is released at about 380 thousand years. First stars form at about 100 to 200 million years, the first galaxies are in place by about 300 million years, accelerated expansion begins at about 8 billion years, and today sits at 13.8 billion years at the right-hand end.

The broad sweep of cosmic history is sketched in Figure 44.9: from the Big Bang through a brief burst of *inflation*, to the release of the cosmic microwave background, to the first stars and galaxies, and finally to the present era of accelerating expansion. The rest of this chapter follows the evidence for that story: starting with the discovery that first told us the universe is expanding at all.

Hubble's Law

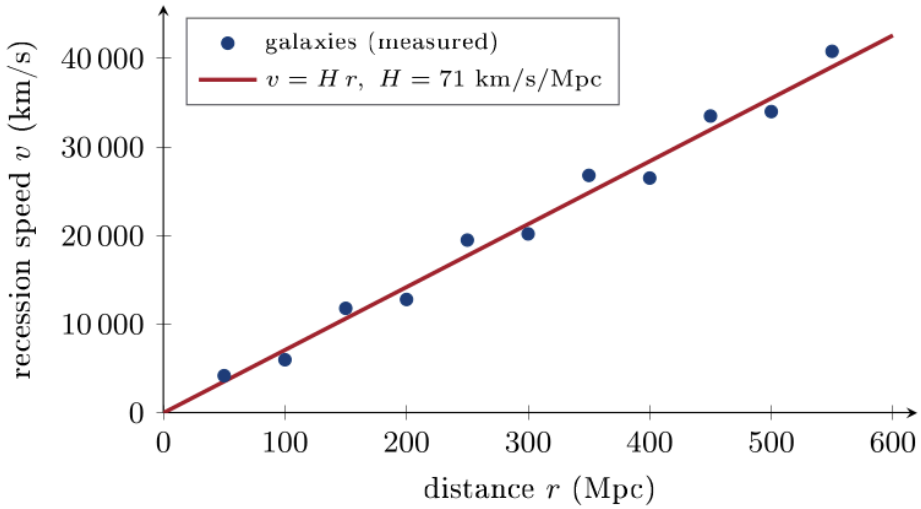


Figure 44.10. Hubble's law. Each point is a galaxy whose recession speed (from its redshift) is plotted against its distance. The data scatter about a straight line through the origin whose slope is the Hubble constant H. The farther a galaxy, the faster it recedes: the signature of a uniformly expanding universe.

•Figure description

A scatter plot of galaxy recession speed in kilometers per second against distance in megaparsecs, out to about 550 Mpc and 41,000 km/s. Eleven measured galaxies are plotted as blue dots. They scatter about a straight red line drawn through the origin with slope 71 kilometers per second per megaparsec, the Hubble constant. The farther the galaxy, the faster it recedes, and the relationship is linear.

The first evidence that the universe is expanding came from Edwin Hubble. Using the Doppler effect (Chapter 35) to read the recession speeds of galaxies from the redshift of their light, Hubble found that almost every galaxy is moving *away* from us, and the farther away a galaxy is, the faster it recedes. The relationship is strikingly simple and linear (Figure 44.10):

Key Equation: Hubble's law

$$v = H\,r, \tag{44.6}$$

where v is a galaxy's recession speed, r its distance, and H the **Hubble constant**.

A linear law is exactly what a uniform expansion produces. If every distance between galaxies grows by the same factor, then in a given time a galaxy twice as far away moves twice as far, so it recedes twice as fast (Figure 44.11). Every galaxy sees the same pattern, so the expansion has no center: we are not in a special place.

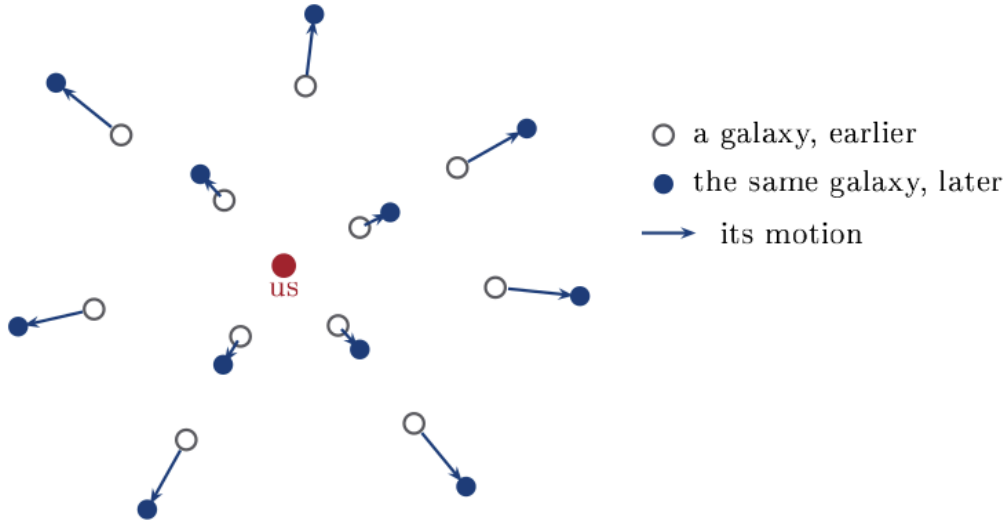


Figure 44.11. Why expansion gives Hubble's law. Every distance grows by the same factor (here 1.4), so in the same time each galaxy moves away from us by 0.4 times its distance: farther galaxies recede faster, $v \propto r$. An observer in any other galaxy would draw the same picture centered on itself.

•Figure description

A sketch of uniform expansion seen from our galaxy, a red dot labeled us at the center. Eleven other galaxies are scattered around it, each drawn twice: as a hollow gray circle where it was earlier and as a filled blue dot where it is later, joined by a blue arrow pointing straight away from us. Every distance has grown by the same factor, so the nearby galaxies have moved only a short way while the distant ones have moved much farther: speed is proportional to distance.

Cosmic distances are measured in megaparsecs, with

$$1\ \text{Mpc} = 3.084 \times 10^{19}\ \text{km} = 3.260 \times 10^{6}\ \text{ly}, \tag{44.4.1}$$

and the measured value of the Hubble constant is

$$H = 21.8 \times 10^{-3}\ \frac{\text{m}}{\text{s}\cdot\text{ly}} = 71.0\ \frac{\text{km}}{\text{s}\cdot\text{Mpc}}. \tag{44.4.2}$$

The Age of the Universe

If every galaxy has been receding at constant speed $v = Hr$, then the time since everything was together is $t = r/v = 1/H$, independent of which galaxy we pick. This *Hubble time* is a first estimate of the age of the universe:

Key Equation: Age of the universe (Hubble time)

$$T \approx \frac{1}{H}. \tag{44.7}$$

With $H = 21.8 \times 10^{-3}\ \text{m}/(\text{s}\cdot\text{ly})$ and $1\ \text{ly} = 9.461 \times 10^{15}\ \text{m}$,

$$T \approx \frac{1}{H} = \frac{9.461 \times 10^{15}\ \text{m}}{21.8 \times 10^{-3}\ \text{m/s}} = 4.34 \times 10^{17}\ \text{s} \approx 13.8 \times 10^{9}\ \text{yr}, \tag{44.4.3}$$

about 13.8 billion years: remarkably close to the age obtained by completely independent methods. (The value of H is itself the subject of an unresolved debate: measurements based on the cosmic microwave background give about $67\ \text{km}/(\text{s}\cdot\text{Mpc})$, while measurements using nearby supernovae give about 73. This disagreement, the *Hubble tension*, may point to physics beyond our current models. We use 71 as a representative value.)

Example 44.5: How far away is a quasar?

A quasar is a brilliant, extremely distant object whose recession speed can be read from the Doppler shift of its light. A certain quasar recedes from Earth at $0.55c$.

a. How far away is it?
b. If it emitted blue light at $\lambda = 450\ \text{nm}$, at what wavelength do we receive it?

Solution

(a) The recession speed is

$$v = 0.55c = 0.55\,(2.998 \times 10^5\ \text{km/s}) = 1.65 \times 10^5\ \text{km/s}. \tag{44.4.4}$$

By Hubble's law, $r = v/H$:

$$r = \frac{1.65 \times 10^5\ \text{km/s}}{71.0\ \text{km/(s}\cdot\text{Mpc)}} = 2.3 \times 10^3\ \text{Mpc}. \tag{44.4.5}$$

Converting to light-years, $r = (2.3 \times 10^3)(3.260 \times 10^6\ \text{ly}) = \boxed{7.6 \times 10^9\ \text{ly}}$: about 7.6 billion light-years.

(b) Using the low-speed Doppler relation from our study of relativity, $v/c = |\Delta\lambda|/\lambda$, the shift is

$$|\Delta\lambda| = \frac{v}{c}\,\lambda = (0.55)(450\ \text{nm}) = 248\ \text{nm}. \tag{44.4.6}$$

The galaxy recedes, so the light is redshifted to a longer wavelength:

$$\lambda_{\text{obs}} = 450 + 248 = \boxed{698\ \text{nm}}, \tag{44.4.7}$$

deep red. The blue light has been stretched all the way across the visible spectrum. (At $0.55c$ the simple formula is only approximate; the fully relativistic Doppler formula gives a somewhat larger shift, near $835\ \text{nm}$, but the qualitative result, a large redshift toward the red, is the same. Likewise, part (a)'s simple $v = Hr$ is a good estimate but not exact at such high speed.)

Cosmic Microwave Background

Hubble's law tells us the universe is expanding; run that expansion backward and the universe must once have been far smaller, denser, and hotter. The strongest evidence that it truly had such a hot, dense beginning is the **cosmic microwave background** (CMB): the relic radiation of the Big Bang itself.

In the first few hundred thousand years the universe was an opaque plasma of free electrons and nuclei; photons could not travel far without scattering. As the universe expanded and cooled to a few thousand kelvin, at an age of about $380{,}000$ years, electrons and nuclei combined into neutral atoms (the era of *recombination*) and the universe abruptly became transparent. The photons set free at that moment have streamed across the cosmos ever since, stretched by the expansion from their original visible/infrared wavelengths all the way into the microwave band. Today they bathe the entire sky as a nearly perfect *blackbody* (Chapter 39) at a temperature of about $2.7\ \text{K}$ (Figure 44.12).

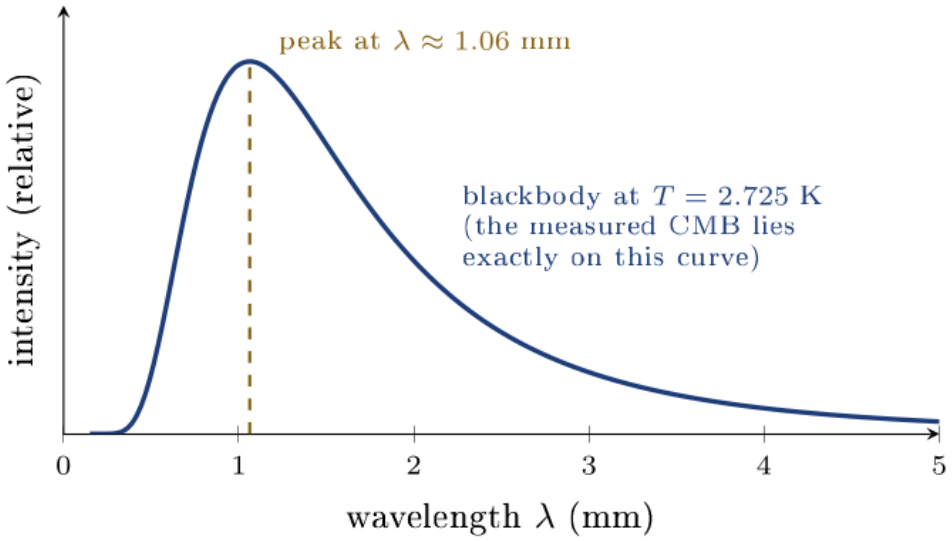


Figure 44.12. The spectrum of the cosmic microwave background: a blackbody at $2.725\ \mathrm{K}$, peaking at a wavelength of about $1\ \mathrm{mm}$, in the microwave band. The measured spectrum matches this curve so closely that the difference would be far thinner than the line.

•Figure description

A graph of the intensity of the cosmic microwave background against wavelength in millimeters, from 0 to 5. The blue curve, a blackbody spectrum at 2.725 kelvin, is near zero below about 0.3 millimeters, rises steeply to a single peak at about 1.06 millimeters, marked by a dashed gold line, and then falls off gradually toward longer wavelengths. A label notes that the measured CMB lies exactly on this curve.

The CMB is not perfectly uniform. It carries faint temperature ripples, of order one part in 10^5, mapped in exquisite detail by satellite observatories. These tiny anisotropies are the seeds from which gravity later grew galaxies and clusters, so the CMB is at once a snapshot of the infant universe and the blueprint of all the structure that followed.

Checkpoint 44.6

The CMB photons were released as visible and infrared light but reach us as microwaves, with wavelengths roughly a thousand times longer. Using Hubble's picture of an expanding universe, explain qualitatively why their wavelengths have grown, and what that stretching implies about the size of the universe then versus now.

Answer

Dark Matter

For all its success, the matter we understand (the quarks and leptons of the Standard Model) turns out to be a small minority of what the universe contains. The first clue came from how galaxies spin.

Galactic Rotation Curves

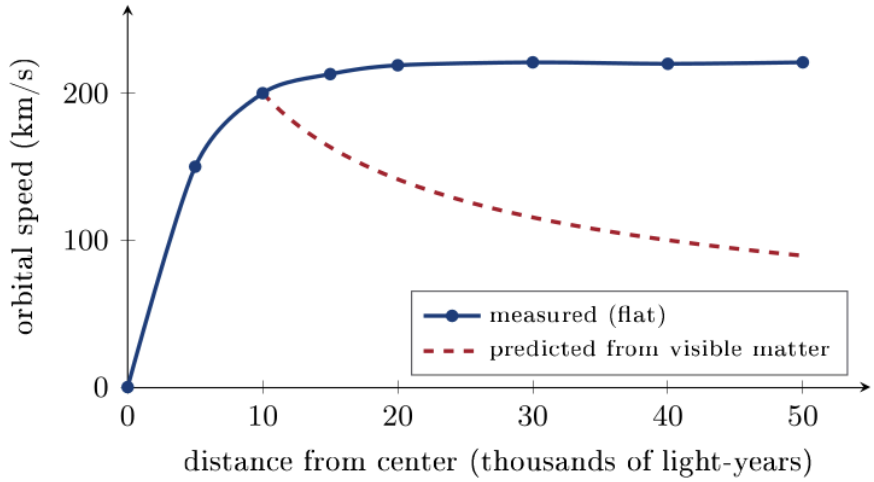


Figure 44.13. A galactic rotation curve. The speed predicted from the visible (luminous) matter falls off with distance, like the orbital speeds of the planets (red dashed). The measured curve instead stays flat far beyond the visible disk (blue), implying a vast amount of unseen mass: *dark matter*.

•Figure description

Two galactic rotation curves plotted as orbital speed against distance from the galaxy's center, out to 50 thousand light-years. Both start at zero and climb together to about 200 kilometers per second by 10 thousand light-years. Beyond that they part company: the red dashed curve, predicted from the visible matter alone, falls steadily away to about 90 kilometers per second at the outer edge, while the blue measured curve stays flat at about 220 all the way out. The gap between the two is the evidence for dark matter.

Gravity weakens with distance, so a star orbiting far from a galaxy's luminous center ought to move *slower* than one nearer in, just as the outer planets orbit the Sun more slowly than the inner ones. The predicted rotation curve (orbital speed versus distance from center) should therefore fall off at large radius. Instead, measurements show the curve staying stubbornly *flat*: stars and gas far out orbit just as fast as those closer in (Figure 44.13). The Andromeda galaxy's rotation rate, for instance, remains roughly constant well beyond its visible edge: something the visible mass alone cannot explain.

We read these curves by measuring the Doppler shift of the 21-cm radiation from cold hydrogen gas, which extends far beyond the visible stars and so lets us clock the orbital speed even where there are few stars to see. The flatness implies that the luminous galaxy sits at the heart of a much larger, invisible halo whose gravity holds the fast outer orbits in place.

Other Evidence

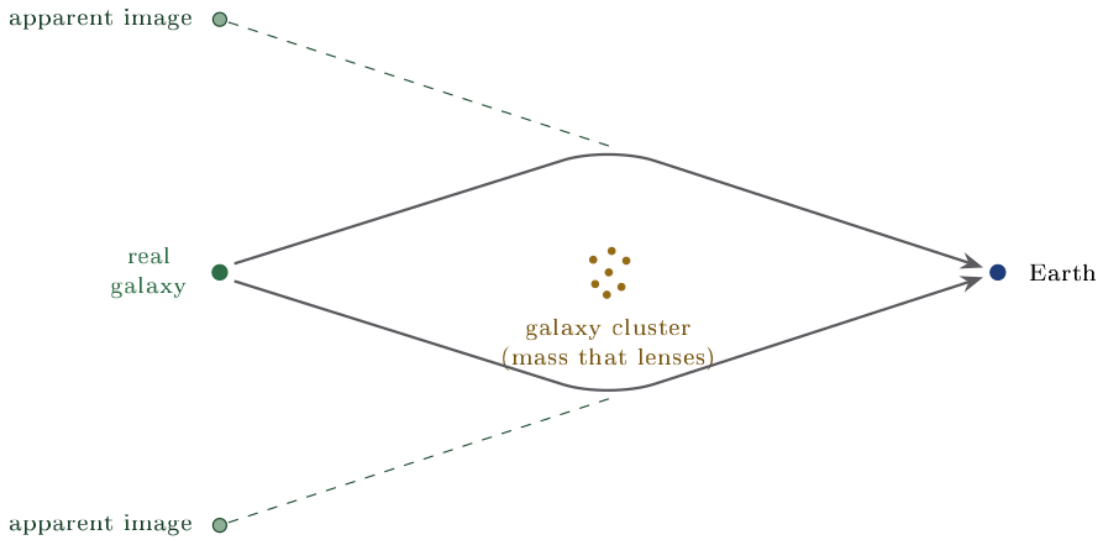


Figure 44.14. Gravitational lensing. The cluster's gravity bends light from a background galaxy so that it reaches Earth along two (or more) curved paths. We see the galaxy displaced into multiple apparent images; the amount of bending measures the cluster's total mass, including its dark matter.

•Figure description

A gravitational lensing diagram. A real galaxy sits at the left, Earth at the right, and a cluster of gold dots, the lensing mass, midway between them. Light from the real galaxy follows two curved gray paths that bend around the cluster, one passing above it and one below, both arriving at Earth. Dashed green lines trace each arriving ray straight backwards from Earth to where the galaxy appears to be; they end at two small green dots labeled apparent image, one above and one below the true position. The observer sees two images of a single galaxy, and the amount of bending measures the cluster's total mass.

Two more lines of evidence point the same way. Doppler measurements of galaxies near the Milky Way reveal a large-scale streaming motion toward a common region of unseen mass nicknamed the *Great Attractor*. And **gravitational lensing** (the bending of light by gravity, Chapter 35) lets us weigh a galaxy cluster directly from how strongly it distorts the images of more distant galaxies behind it (Figure 44.14). Cluster after cluster turns out to be far more massive than its visible matter can account for.

What Is It?

The conclusion is that most of a galaxy's mass lies in an invisible halo far larger than the galaxy itself: **dark matter**. What it is made of remains unknown; there has been no direct detection of a dark-matter interaction. Searches for *MACHO*s (massive compact halo objects, made of ordinary matter that simply does not shine) have found far too few to account for it, so most physicists expect a new kind of particle outside the Standard Model. Leading candidates include *WIMP*s (weakly interacting massive particles) and *axions* (extremely light particles), but decades of increasingly sensitive detectors have yet to see either. Identifying dark matter is one of the central open problems in physics.

Dark Energy

If dark matter were the whole story, gravity would gradually slow the cosmic expansion. But careful measurements of very distant supernovae reveal the opposite: the expansion is *speeding up*. The universe is accelerating outward.

This is deeply strange. Acceleration requires a net outward push, yet the gravity of all the matter (visible and dark) should be *decelerating* the expansion. Something with a kind of negative pressure must be driving space apart, and we do not know what it is. We give the unknown cause a name (**dark energy**) but naming is not explaining.

Putting the pieces together yields a sobering inventory of the cosmos. Ordinary Standard-Model matter (everything in this book up to this chapter) makes up only about 5% of the universe's total energy content. Dark matter contributes roughly 27%, and dark energy the remaining $\sim 68\%$. In other words, about 95% of the universe is of a nature we have not yet identified. The Standard

Model, the crowning achievement of twentieth-century physics, describes only the small visible sliver.

Checkpoint 44.7

A friend says, “Gravity is attractive, so the expansion of the universe must be slowing down.” Given the supernova evidence, what is wrong with this reasoning, and what does the correction imply about the dominant contribution to the universe’s energy budget?

Answer

The Ultimate Fate of the Universe

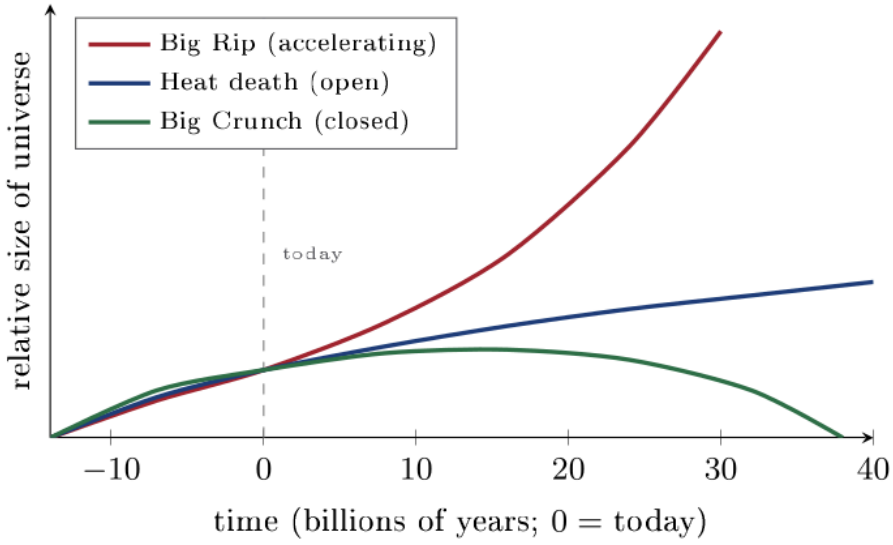


Figure 44.15. Three possible histories of cosmic expansion, plotting the relative size of the universe against time (today at $t = 0$). Whether the universe recollapses (Big Crunch), expands forever toward heat death, or is torn apart (Big Rip) depends on the balance between gravitating matter and dark energy.

•Figure description

A graph of the relative size of the universe against time in billions of years, with today marked at zero by a dashed vertical line. All three curves rise together from the Big Bang on the left and pass through the same point today, then diverge. The red Big Rip curve sweeps upward ever more steeply. The blue heat-death curve keeps rising but flattens gradually, coasting outward forever. The green Big Crunch curve levels off, turns over, and falls back to zero at about 38 billion years.

How the universe ends depends on the contest between gravity, pulling everything together, and dark energy, pushing everything apart, that is, on its total content of matter and energy. Three broad fates are possible (Figure 44.15):

- **Big Crunch (or Big Bounce).** If gravity ultimately wins, the expansion halts and reverses; the universe collapses back to a hot, dense state. A “bounce” would have it re-expand into a new cycle.
- **Heat Death.** If the universe expands forever at a roughly steady or gently slowing rate, it cools toward uniformity. Stars burn out, matter disperses, and the universe drifts to a state of maximum entropy in which no further work can be done, the thermodynamic “heat death” foreshadowed by the second law (Chapter 38).
- **Big Rip.** If dark energy not only persists but *strengthens*, the acceleration runs away: galaxies, then stars, then atoms, and finally space itself are torn apart.

Current measurements (a universe dominated by dark energy and expanding ever faster) favor an open future closer to heat death (or, if dark energy grows, a Big Rip) than to a Big Crunch. But the answer hinges on the true nature of dark energy, which remains unknown: in 2025 the Dark Energy Spectroscopic Instrument (DESI) found hints that dark energy may be slowly weakening, which, if confirmed, would change the forecast. We end this book, fittingly, at the frontier: the same conservation laws and the same second law of thermodynamics that we built up over the course of this book still hold sway over the entire universe, even as its deepest constituents and its ultimate fate remain open questions for the next generation of physicists, perhaps including you.

Answers to Checkpoints

Checkpoint 44.6. As the universe expands, space itself stretches, and the wavelength of light traveling through it stretches by the same factor. The CMB's wavelengths have grown about a thousandfold since it was released, so the universe has expanded by that same factor: distances between widely separated points were about a thousand times smaller then than they are now, and the universe was correspondingly hotter and denser.

Checkpoint 44.7. The reasoning would be right if ordinary gravity were all there is. But the supernova observations show that distant supernovae are dimmer (farther away) than a slowing expansion predicts: the expansion is s*peeding up*. Something must drive that acceleration, and it cannot be ordinary matter, whose gravity only attracts. The correction points to *dark energy*, which fills space and pushes the expansion outward. To produce the observed acceleration it must be the largest part of the universe's energy budget, about 70%, outweighing all matter, dark and ordinary, combined.

This page titled 44.4: Cosmology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.S: Particle Physics and Cosmology (Summary)

Summary

- Particles are classified by **spin** into *fermions* (half-integer spin, obey the Pauli principle: quarks, leptons) and *bosons* (integer spin, force carriers); and by the strong force into *hadrons* (feel it; built from quarks) and *leptons* (do not).
- **Hadrons** split into *baryons* (three quarks, $B = +1$, fermions) and *mesons* ($q\bar{q}$, $B = 0$, bosons). Every particle has an **antiparticle** with opposite charge-like quantum numbers; particle–antiparticle pairs annihilate, releasing $\geq 2mc^2$.
- Reactions conserve charge, **baryon number**, **lepton number** (per family), and **strangeness** (the last only in strong/EM interactions). Use these to test whether a reaction can occur.
- The **Standard Model** has six quarks, six leptons, the gauge bosons (gluon, photon, W, Z), and the **Higgs**. The four forces are gravity, electromagnetic, weak, and strong; EM and weak are unified as the *electroweak* force. The Higgs field gives the W and Z their mass.
- Collider **detectors** identify particles by their signatures in the tracker, calorimeters, and muon chambers. A track's curvature gives its momentum, $pc = (0.2998\ \text{GeV})(B/1\ \text{T})(r/1\ \text{m})$; neutrinos show up as *missing transverse momentum*; and the **invariant mass** of the decay products, $(Mc^2)^2 = (\sum E)^2 - |\sum \mathbf{p}\,c|^2$, reveals a short-lived particle such as the Higgs as a bump at its mass.
- **Hubble's law** $v = Hr$ shows the universe is expanding; $H = 71.0\ \text{km/(s}\cdot\text{Mpc)}$ gives an age $T \approx 1/H \approx 13.8$ billion years.
- The **cosmic microwave background** (a ~ 2.7 K blackbody) is relic light from recombination ($\sim 380{,}000$ yr), strong evidence for a hot Big Bang.
- **Dark matter** (flat rotation curves, lensing) and **dark energy** (accelerating expansion) make up $\sim 95\%$ of the universe; ordinary matter is only $\sim 5\%$. The universe's fate (Big Crunch, heat death, or Big Rip) depends on the balance between them.

Key Equations

Quantized spin component

$$S_z = m_s\,\hbar, \qquad m_s = -s,\ -s+1,\ \ldots,\ +s \tag{44.S.1}$$

Fermions and bosons

$$\begin{aligned} \text{fermions:}\quad & s = \tfrac{1}{2}, \tfrac{3}{2}, \ldots \quad && \text{obey the Pauli exclusion principle} \\ \text{bosons:}\quad & s = 0, 1, 2, \ldots && \text{no such restriction} \end{aligned} \tag{44.S.2}$$

Particle–antiparticle annihilation energy

$$E_{\text{annih}} = 2mc^2 \tag{44.S.3}$$

Conserved in every reaction

$$q, \quad B, \quad L_e,\ L_\mu,\ L_\tau, \quad S\ \text{(strong and EM only)} \tag{44.S.4}$$

Baryon and meson quark content

$$\begin{aligned} \text{baryon:}\quad & qqq, && B = 3 \times \tfrac{1}{3} = +1 \\ \text{meson:}\quad & q\bar{q}, && B = \tfrac{1}{3} - \tfrac{1}{3} = 0 \end{aligned} \tag{44.S.5}$$

Momentum from track curvature

$$pc = (0.2998\,\text{GeV})\left(\frac{B}{1\,\text{T}}\right)\left(\frac{r}{1\,\text{m}}\right) \tag{44.S.6}$$

Invariant mass of decay products

$$(Mc^2)^2 = \left(\sum E\right)^2 - \left|\sum \mathbf{p}\,c\right|^2 \qquad \text{(two photons: } (Mc^2)^2 = 2E_1E_2(1-\cos\theta)\text{)} \tag{44.S.7}$$

Hubble's law

$$v = H\,r, \qquad H = 71.0\ \frac{\text{km}}{\text{s}\cdot\text{Mpc}} = 21.8\times10^{-3}\ \frac{\text{m}}{\text{s}\cdot\text{ly}} \tag{44.S.8}$$

Age of the universe (Hubble time)

$$T \approx \frac{1}{H} \approx 13.8\times10^{9}\ \text{yr} \tag{44.S.9}$$

Energy content of the universe

$$\underbrace{\sim 5\%}_{\text{ordinary matter}} + \underbrace{\sim 27\%}_{\text{dark matter}} + \underbrace{\sim 68\%}_{\text{dark energy}} \tag{44.S.10}$$

This page titled 44.S: Particle Physics and Cosmology (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

44.E: Particle Physics and Cosmology (Problems)

Difficulty key: ★ introductory, ★★ standard, ★★★ challenging. Useful data: $c = 3.00 \times 10^8\ \text{m/s}$; $m_\pi c^2 = 139.6\ \text{MeV}$, $m_\mu c^2 = 105.7\ \text{MeV}$, $m_e c^2 = 0.511\ \text{MeV}$; $H = 71.0\ \text{km/(s}\cdot\text{Mpc)}$; $1\ \text{Mpc} = 3.084 \times 10^{19}\ \text{km} = 3.260 \times 10^6\ \text{ly}$; $1\ \text{ly} = 9.461 \times 10^{15}\ \text{m}$. Quark quantum numbers from Table 44.1.

Sections 44.1–44.1.2 — Classifying Particles

44.1 ★ *[Conceptual.]* For each particle, state whether it is a fermion or a boson, and whether it is a hadron or a lepton: (a) the electron; (b) the proton; (c) the photon; (d) the π^+ meson; (e) the electron neutrino.

44.2 ★ A particle has spin $s = \frac{3}{2}$. (a) Is it a fermion or a boson? (b) List the allowed values of m_s. (c) What is the largest value of S_z, in units of $\hbar$?

44.3 ★★ *[Conceptual.]* Explain, using the Pauli exclusion principle, why a Bose–Einstein condensate can form from bosonic atoms but not from a gas of identical fermions.

44.4 ★ *[Real-world application: ultracold atoms.]* A neutral atom is a fermion or a boson according to whether its total number of spin-$\frac{1}{2}$ constituents (protons, neutrons, and electrons) is odd or even.

a. Classify each neutral atom: ^3He, ^4He, ^6Li, ^7Li, and ^{87}Rb.
b. Show that for a neutral atom only the number of neutrons decides the answer.
c. The first Bose–Einstein condensate (1995) was made from one of these five kinds of atom. Which of them could have been used?

44.5 ★★ The magnitude of a particle's spin angular momentum is $|\mathbf{S}| = \sqrt{s(s+1)}\ \hbar$. (a) Find $|\mathbf{S}|$, in J·s, for an electron ($s = \frac{1}{2}$) and for a photon ($s = 1$). (b) For an electron with $m_s = +\frac{1}{2}$, find the angle between $\mathbf{S}$ and the z-axis. (c) Explain why $\mathbf{S}$ can never point exactly along the z-axis.

Sections 44.1.3–44.1.4 — Antiparticles and Quarks

44.6 ★ How much energy (in MeV) is released when an electron and a positron at rest annihilate completely? If the energy emerges as two identical photons, what is the energy of each?

44.7 ★★ Determine the charge, strangeness, and baryon number of each quark combination, and identify whether it is a baryon or a meson: (a) uud; (b) $u\bar{s}$; (c) sss; (d) $u\bar{d}$.

44.8 ★★ The Σ^+ baryon has charge $q = +1$, strangeness $S = -1$, and baryon number $B = +1$. Find its quark content. *(Hint: it contains exactly one strange quark.)*

44.9 ★★ Use conservation laws to decide whether each reaction can occur, and name the violated quantity if not: (a) $p \to n + e^+ + \nu_e$; (b) $p + \bar{p} \to \gamma + \gamma$; (c) $n \to p + e^-$ (with no other particles).

44.10 ★★★ A stationary K^+ meson decays by $K^+ \to \mu^+ + \nu_\mu$. Given $m_K c^2 = 493.7\ \text{MeV}$ and $m_\mu c^2 = 105.7\ \text{MeV}$ (treat the neutrino as massless), find the kinetic energy of the muon and of the neutrino. *(Follow the method of the pion-decay example.)*

44.11 ★★ *[Real-world application: discovering particles.]* In each reaction, use conservation of charge, baryon number, and lepton number, and for the strong reactions strangeness, to find the quantum numbers of the unknown particle X, and identify it.

a. $\bar{\nu}_e + p \to n + X$, the reaction used in 1956 to detect the neutrino for the first time.
b. $\pi^- + p \to K^0 + X$, a strong reaction ($K^0 = d\bar{s}$).
c. $K^- + p \to \Omega^- + K^+ + X$, the strong reaction that discovered the $\Omega^- = sss$ in 1964 ($K^- = s\bar{u}$, $K^+ = u\bar{s}$).

44.12 ★★ The antiproton $\bar{p}$ is the antiparticle of the proton (uud). (a) Give its quark content. (b) State its electric charge and baryon number. (c) Find the energy released when a proton and antiproton, both at rest, annihilate. (Take the proton rest energy

as $938\ \mathrm{MeV}$.)

Sections 44.2–44.3 — Forces and the Standard Model

44.13 ★ *[Reading Feynman diagrams.]* Electric charge is conserved at every vertex of a Feynman diagram: the total charge of the lines entering a vertex equals the total charge leaving it.

a. In the beta-decay diagram (Figure 44.5), a d quark becomes a u quark and emits the W. Use charge conservation at that vertex to find the charge of the W.
b. Check that charge is also conserved at the second vertex, where the W becomes an electron and an antineutrino.
c. In the electron–electron diagram (Figure 44.4), what is the charge of the exchanged photon, and why must it be so?

44.14 ★★ *[Conceptual.]* The strong force binds quarks into color-neutral hadrons. Show that the combinations RGB (three quarks) and $R\bar{R}$ (quark–antiquark) are both color-neutral, and explain why a free single quark is never observed.

44.15 ★★ *[Order-of-magnitude estimate.]* A messenger particle of mass m can be "borrowed" only briefly: by the energy–time uncertainty relation (Chapter 39), an energy $\Delta E = mc^2$ can exist for about $\Delta t \approx \hbar/\Delta E$. Even moving at nearly the speed of light, the carrier then travels only about $R \approx c\ \Delta t$ before it must be reabsorbed, and R sets the range of the force. (a) Estimate Δt and R for the W boson, whose rest energy is $80.4\ \mathrm{GeV}$ (use $\hbar = 6.582 \times 10^{-16}\ \mathrm{eV \cdot s}$). (b) Compare R with the weak-force range in Table 44.2 and with the size of a proton, about $1 \times 10^{-15}\ \mathrm{m}$. (c) Use the same reasoning to explain why the photon gives the electromagnetic force an unlimited range.

44.16 ★★ *[Estimation: when the forces were one.]* Forces that are unified at high energy separate once the typical particle energy $k_B T$ falls below the unification energy. In the early universe the temperature fell with time roughly as $T \approx (1.5 \times 10^{10}\ \mathrm{K})/\sqrt{t/(1\,\mathrm{s})}$. Use $k_B = 8.617 \times 10^{-5}\ \mathrm{eV/K}$.

a. Electroweak unification sets in near $100\ \mathrm{GeV}$. Find the temperature at which $k_B T = 100\ \mathrm{GeV}$, and how long after the Big Bang the universe cooled through it.
b. Grand unification is expected near $10^{15}\ \mathrm{GeV}$. Repeat (a).
c. The Large Hadron Collider reaches collision energies of about $10^{4}\ \mathrm{GeV}$. Explain why electroweak unification can be tested directly in the laboratory while grand unification can be tested only indirectly.

Section 44.3.1 — Detecting Particles

44.17 ★ *[Conceptual.]* Identify the particle that leaves each signature in a collider detector: (a) a curved track that ends in the electromagnetic calorimeter; (b) no track, but a deposit in the electromagnetic calorimeter; (c) a curved track that ends in the hadron calorimeter; (d) no track, but a deposit in the hadron calorimeter; (e) a nearly straight track that continues through the muon chambers; (f) nothing at all, except that the visible momenta across the beam do not add to zero.

44.18 ★★ *[Real-world application: the CMS tracker.]* The CMS tracker sits in a $3.8\ \mathrm{T}$ field and follows a charged particle out to a radius of about $L = 1.1\ \mathrm{m}$.

a. Two singly charged tracks have radii of curvature $1.8\ \mathrm{m}$ and $18\ \mathrm{m}$. Find the momentum of each, as pc in GeV.
b. Over a length L, a circular arc of radius $r \gg L$ departs from a straight line by its *sagitta*, $s \approx L^2/8r$. Find s for each track.
c. Find the radius of curvature and the sagitta of a $1.0\ \mathrm{TeV}$ muon. The tracker locates each point to about $10\ \mu\mathrm{m}$. Explain why momentum is harder to measure the more energetic the particle is, and why measuring muons again in chambers several meters out helps.

44.19 ★★ *[Real-world application: rediscovering the Z.]* A Z boson can decay to a muon and an antimuon. Muons are so light ($m_\mu c^2 = 0.106\ \mathrm{GeV}$) that at these energies you may treat them as massless, as for photons.

a. A Z at rest decays, sending the two muons off back to back with $45.6\ \mathrm{GeV}$ each. Find the invariant mass of the pair.
b. In another event the muons have energies of $60.0\ \mathrm{GeV}$ and $40.0\ \mathrm{GeV}$, with 137° between them. Could this pair also have come from a Z?
c. Why must the muons in (a) fly apart back to back?

44.20 ★★ *[Missing momentum.]* In the plane perpendicular to the beam, a detector records an electron with transverse momentum $40.0\ \mathrm{GeV/c}$ along the $+x$ axis and a jet with $15.0\ \mathrm{GeV/c}$ at 120° from the $+x$ axis. Nothing else is seen. Find the

magnitude and direction of the missing transverse momentum, and suggest what carried it away.

Section 44.4 — Cosmology, Dark Matter, and the Fate of the Universe

44.21 ★ A galaxy is $150\ \text{Mpc}$ away. Using Hubble's law, find its recession speed in km/s.

44.22 ★★ *[Real-world application: astronomy.]* A galaxy recedes at $2.5 \times 10^4\ \text{km/s}$. (a) Find its distance in Mpc and in light-years. (b) Roughly how long ago did the light we now see leave it?

44.23 ★★ *[Conceptual.]* Sketch a galactic rotation curve as predicted from visible matter and as actually measured. Explain what the difference between the two curves tells us, and name two independent observations that support the same conclusion.

44.24 ★★★ *[Capstone — back to the night sky.]* Recall the chapter opener: the ancient light of a distant quasar, and the question of what the universe is made of and how old it is.

a. Estimate the age of the universe from the Hubble constant, $T \approx 1/H$, expressing your answer in years. *(Use* $H = 21.8 \times 10^{-3}\ \text{m/(s}\cdot\text{ly)}$ *and* $1\ \text{ly} = 9.461 \times 10^{15}\ \text{m}$*.)*
b. The quasar in the chapter recedes at $0.55c$. Using $v = Hr$, find its distance in light-years, and comment on how it compares with the age in (a), i.e. roughly how far back in cosmic history its light was emitted.
c. Of the total energy content of the universe, ordinary (Standard-Model) matter is about 5%, dark matter about 27%, and dark energy about 68%. If the quasar, its host galaxy, and everything in this book are made of that 5%, what fraction of the universe's content remains unidentified? In one or two sentences, state what dark matter and dark energy are each inferred from, and why naming them is not the same as understanding them.

Challenge Problems

44.25 ★★★ *[Real-world application: particle physics — why we build colliders.]* What matters for creating new particles is not a beam's energy but the energy available in the center-of-mass frame, $\sqrt{s}$. For two particles of rest energy mc^2, the invariant s satisfies $s = (E_1 + E_2)^2 - (p_1c + p_2c)^2$, and for highly relativistic particles $pc \approx E$.

a. Two protons of energy E each collide head-on. Show that $\sqrt{s} = 2E$, and evaluate it for the LHC's $E = 6.5\ \text{TeV}$ beams.
b. A beam of energy E strikes a proton at rest. Show that $\sqrt{s} = \sqrt{2mc^2E + 2(mc^2)^2}$, and evaluate it for $E = 6.5\ \text{TeV}$, using $mc^2 = 938.272\ \text{MeV}$.
c. What beam energy would a fixed-target machine need in order to match the LHC's collider $\sqrt{s}$? Express it in TeV.
d. Compare your answers, and explain why colliders (despite the enormous difficulty of steering two beams into each other) replaced fixed-target machines at the energy frontier. Note also one genuine advantage a fixed-target experiment retains.

44.26 ★★★ *[Real-world application: cosmology — weighing the Milky Way.]* A star in a circular orbit at radius r about the center of a galaxy obeys $v^2/r = GM(<r)/r^2$, where $M(<r)$ is the mass enclosed within its orbit. The Sun orbits at $r = 8.2\ \text{kpc}$ with $v = 235\ \text{km/s}$; at $r = 40\ \text{kpc}$ the rotation curve is still essentially flat, at $v = 220\ \text{km/s}$. Take $1\ \text{kpc} = 3.086 \times 10^{19}\ \text{m}$ and $M_\odot = 1.99 \times 10^{30}\ \text{kg}$.

a. Find the mass enclosed within the Sun's orbit, in kilograms and in solar masses.
b. Find the mass enclosed within $40\ \text{kpc}$.
c. Essentially all of the Milky Way's *visible* matter (stars, gas, dust) lies within about $15\ \text{kpc}$ and totals roughly 6×10^{10} solar masses. What fraction of the mass inside $40\ \text{kpc}$ is therefore invisible?
d. If the galaxy's mass really ended at the edge of the visible disk, the rotation curve beyond it would be Keplerian, $v \propto r^{-1/2}$. Predict the orbital speed at $40\ \text{kpc}$ on that assumption, compare it with the measured $220\ \text{km/s}$, and state plainly what the discrepancy is evidence for.

44.27 ★★★ *[Real-world application: cosmology — reading the oldest light.]* The cosmic microwave background is a nearly perfect blackbody at $T = 2.725\ \text{K}$. It was released when the universe had cooled enough for atoms to form, at about $3000\ \text{K}$. Use Wien's law, $\lambda_{\max}T = 2.898 \times 10^{-3}\ \text{m}\cdot\text{K}$, and the chapter's $H = 21.8 \times 10^{-3}\ \text{m/(s}\cdot\text{ly)}$.

a. Find the wavelength at which the CMB spectrum peaks today, and confirm that it lies in the microwave band. Find the energy of such a photon, in eV, and compare it with the $13.6\ \text{eV}$ needed to ionize hydrogen.
b. Wavelengths stretch in proportion to the expansion of the universe, so $\lambda \propto 1/T$. By what factor has the universe expanded

since the CMB was released? What redshift $z = \Delta\lambda/\lambda_{\text{emitted}}$ does that correspond to?

c. A quasar's spectrum is redshifted so that it recedes at $0.10c$. Use Hubble's law to find its distance in light-years, and compare it with the Hubble time of 13.8×10^9 yr.

d. Part (a) shows that CMB photons today are far too feeble to ionize anything, while at 3000 K they were not. Explain in two or three sentences why the universe was opaque before that moment and transparent after it, and why this makes the CMB the oldest light anyone will ever see.

This page titled 44.E: Particle Physics and Cosmology (Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

A: Physical Constants and Astronomical Data

The values below are rounded to the precision typically needed for the problems in this book. The free-fall acceleration g varies slightly with location and altitude; the standard value is used throughout.

Useful Physical Constants

Useful physical constants

Quantity	Symbol	Value
Speed of light in vacuum	c	$2.998 \times 10^{8}\ \text{m/s}$
Gravitational constant	G	$6.674 \times 10^{-11}\ \text{N}\cdot\text{m}^2/\text{kg}^2$
Standard free-fall acceleration	g	$9.80\ \text{m/s}^2$
Standard atmospheric pressure	P_0	$1.013 \times 10^{5}\ \text{Pa}$
Speed of sound in air (20°C)	v_{sound}	$343\ \text{m/s}$
Density of water (4°C)	ρ_w	$1.000 \times 10^{3}\ \text{kg/m}^3$
Density of air (sea level, 20°C)	ρ_{air}	$1.20\ \text{kg/m}^3$
Density of seawater	ρ_{sea}	$1.03 \times 10^{3}\ \text{kg/m}^3$
Avogadro's number	N_A	$6.022 \times 10^{23}\ \text{mol}^{-1}$

Astronomical and Earth Data

Astronomical and Earth data

Body / Quantity	Symbol	Value
Mass of the Earth	M_E	$5.97 \times 10^{24}\ \text{kg}$
Mean radius of the Earth	R_E	$6.37 \times 10^{6}\ \text{m}$
Mass of the Moon	M_{Moon}	$7.35 \times 10^{22}\ \text{kg}$
Radius of the Moon	R_{Moon}	$1.74 \times 10^{6}\ \text{m}$
Earth–Moon distance (mean)	d_{EM}	$3.84 \times 10^{8}\ \text{m}$
Orbital period of the Moon	T_{Moon}	$2.36 \times 10^{6}\ \text{s}$ (27.3 d)
Mass of the Sun	$M_\odot$	$1.99 \times 10^{30}\ \text{kg}$
Radius of the Sun	$R_\odot$	$6.96 \times 10^{8}\ \text{m}$
Earth–Sun distance (1 AU)	d_{ES}	$1.50 \times 10^{11}\ \text{m}$
Orbital period of the Earth	T_E	$3.156 \times 10^{7}\ \text{s}$ (365.25 d)

This page titled A: Physical Constants and Astronomical Data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

B: SI Units, Prefixes, and Conversion Factors

SI Base Units

SI base units

Quantity	Unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

Derived Units Used in Mechanics

Derived units used in mechanics

Quantity	Unit	Symbol	In base units
Force	newton	N	$\mathrm{kg \cdot m/s^2}$
Energy, work	joule	J	$\mathrm{kg \cdot m^2/s^2}$
Power	watt	W	$\mathrm{kg \cdot m^2/s^3}$
Pressure	pascal	Pa	$\mathrm{kg/(m \cdot s^2)}$
Frequency	hertz	Hz	$\mathrm{s^{-1}}$
Plane angle	radian	rad	dimensionless

SI Prefixes

SI prefixes

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{24}	yotta-	Y	10^{-1}	deci-	d
10^{21}	zetta-	Z	10^{-2}	centi-	c
10^{18}	exa-	E	10^{-3}	milli-	m
10^{15}	peta-	P	10^{-6}	micro-	μ
10^{12}	tera-	T	10^{-9}	nano-	n
10^{9}	giga-	G	10^{-12}	pico-	p
10^{6}	mega-	M	10^{-15}	femto-	f
10^{3}	kilo-	k	10^{-18}	atto-	a
10^{2}	hecto-	h	10^{-21}	zepto-	z
10^{1}	deka-	da	10^{-24}	yocto-	y

Conversion Factors

Conversion factors

Quantity	Conversion
Length	$1\ \text{in} = 2.54\ \text{cm};\quad 1\ \text{ft} = 0.3048\ \text{m};\quad 1\ \text{mi} = 1.609\ \text{km}$
Mass	$1\ \text{slug} = 14.59\ \text{kg};\quad 1\ \text{u} = 1.661 \times 10^{-27}\ \text{kg}$
Time	$1\ \text{h} = 3600\ \text{s};\quad 1\ \text{d} = 8.64 \times 10^{4}\ \text{s};\quad 1\ \text{yr} = 3.156 \times 10^{7}\ \text{s}$
Speed	$1\ \text{km/h} = 0.278\ \text{m/s};\quad 1\ \text{mi/h} = 0.447\ \text{m/s}$
Force	$1\ \text{lb} = 4.448\ \text{N}$
Energy	$1\ \text{cal} = 4.186\ \text{J};\quad 1\ \text{eV} = 1.602 \times 10^{-19}\ \text{J};\quad 1\ \text{kWh} = 3.60 \times 10^{6}\ \text{J}$
Power	$1\ \text{hp} = 746\ \text{W}$
Pressure	$1\ \text{atm} = 1.013 \times 10^{5}\ \text{Pa} = 760\ \text{mmHg}$
Angle	$1\ \text{rad} = 57.30^{\circ};\quad 180^{\circ} = \pi\ \text{rad}$

The Greek Alphabet

The Greek alphabet

Name	Lower	Upper		Name	Lower	Upper
Alpha	α	A		Nu	ν	N
Beta	β	B		Xi	ξ	Ξ
Gamma	γ	Γ		Omicron	o	O
Delta	δ	Δ		Pi	π	Π
Epsilon	ϵ	E		Rho	ρ	P
Zeta	ζ	Z		Sigma	σ	Σ
Eta	η	H		Tau	τ	T
Theta	θ	Θ		Upsilon	υ	Υ
Iota	ι	I		Phi	ϕ	Φ
Kappa	κ	K		Chi	χ	X
Lambda	λ	Λ		Psi	ψ	Ψ
Mu	μ	M		Omega	ω	Ω

This page titled B: SI Units, Prefixes, and Conversion Factors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C: BASIC ALGEBRA AND TRIGONOMETRY FOR PHYSICS STUDENTS

Most of the mathematics in a physics problem is not calculus. It is algebra: rearranging an equation to isolate the quantity you want, solving a quadratic for a time, or combining two or three equations to find two or three unknowns. And it is trigonometry: splitting a force or a velocity into x- and y-parts, putting the parts back together, or keeping track of the timing of things that oscillate. These skills come up on nearly every page of this book, and a small slip in any of them can sink an otherwise correct solution. This appendix collects them in one place, with physics examples, so you can refer back whenever a step in a chapter feels shaky.

Learning Objectives

- Solve an equation for any one of its symbols, and check the result with units and by substitution.
- Solve a quadratic equation, both when one root is zero and when neither is, and decide which root is physical.
- Solve for two unknowns with two equations, by substitution or by elimination.
- Solve for three unknowns with three equations: reduce them to two, solve, and back-substitute.
- Use sine, cosine, tangent and the Pythagorean theorem to solve a right triangle.
- Break a vector into x- and y-components, and rebuild its magnitude and direction from its components.
- Describe a sinusoid by its amplitude, angular frequency, and phase; represent it by a phasor; and add two sinusoids of the same frequency by adding their phasors.

Note: Solve with symbols first

Physicists almost always solve an equation *symbolically* first and substitute numbers only at the end. The symbolic answer shows how the result depends on each quantity, lets you check its units, and can be reused for new numbers without redoing the algebra. It also keeps rounding errors from piling up.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

C.1: Solving an Equation for One Unknown
C.2: Quadratic Equations
C.3: Solving for Two Unknowns
C.4: Solving for Three Unknowns
C.5: Right-Triangle Trigonometry
C.6: Breaking a Vector into Components
C.7: Finding the Hypotenuse from the Components
C.8: Sinusoids and Phasors
C.S: Basic Algebra and Trigonometry for Physics Students (Summary)
C.E: Basic Algebra and Trigonometry for Physics Students (Practice Problems)

This page titled C: Basic Algebra and Trigonometry for Physics Students is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.1: Solving an Equation for One Unknown

An equation is a balance: the left side equals the right side. You may do anything you like to it, provided you do *the same thing to both sides*: add or subtract the same quantity, multiply or divide by the same nonzero quantity, square both sides, take the square root of both sides. The goal is to end with the unknown alone on one side.

A useful way to plan the steps is to ask what was done to the unknown, in what order, and then undo those operations in *reverse* order. Undo additions and subtractions first, then multiplications and divisions, then powers and roots.

Example C.1: Stopping distance, forward and backward

The kinematic equation $v^2 = v_0^2 + 2a\ \Delta x$ relates speeds, acceleration and displacement. (a) A car traveling at $25\ \mathrm{m/s}$ brakes with acceleration $a = -5.0\ \mathrm{m/s^2}$. How far does it travel before stopping? (b) A different car leaves skid marks $40\ \mathrm{m}$ long, braking at $a = -6.0\ \mathrm{m/s^2}$. How fast was it going when the brakes locked?

Solution

(a) The unknown is Δx. It is multiplied by $2a$, and then v_0^2 is added. Undo in reverse order: subtract v_0^2 from both sides, then divide both sides by $2a$:

$$v^2 - v_0^2 = 2a\,\Delta x \quad\Longrightarrow\quad \Delta x = \frac{v^2 - v_0^2}{2a} = \frac{0 - (25\,\mathrm{m/s})^2}{2(-5.0\,\mathrm{m/s^2})} = 62.5\,\mathrm{m}. \tag{C.1.1}$$

Units check: $(\ \mathrm{m^2/s^2})/(\ \mathrm{m/s^2}) =\ \mathrm{m}$. The two minus signs cancel, giving a positive distance, as they should.

(b) Now the unknown is v_0. Subtract $2a\ \Delta x$ from both sides, then take the square root:

$$v_0 = \sqrt{v^2 - 2a\,\Delta x} = \sqrt{0 - 2(-6.0\,\mathrm{m/s^2})(40\,\mathrm{m})} = \sqrt{480\,\mathrm{m^2/s^2}} = 22\,\mathrm{m/s}. \tag{C.1.2}$$

A square root really has two answers, $\pm$. Here we keep the positive one because we want a speed.

Three common traps

- **A fraction over a sum does not split the other way.** You may split $\dfrac{a+b}{c} = \dfrac{a}{c} + \dfrac{b}{c}$, but $\dfrac{c}{a+b}$ is *not* $\dfrac{c}{a} + \dfrac{c}{b}$.
- **A sum of reciprocals is not the reciprocal of a sum.** If $\dfrac{1}{R} = \dfrac{1}{R_1} + \dfrac{1}{R_2}$, then R is *not* $R_1 + R_2$. Combine the right side over a common denominator first, then flip both sides (Example C.2).
- **Never divide by something that might be zero.** Dividing both sides by the unknown can throw away a perfectly good solution, $x = 0$. Section C.2 shows how to avoid this.

Example C.2: A missing resistor

Two resistors in parallel have an equivalent resistance R given by $\dfrac{1}{R} = \dfrac{1}{R_1} + \dfrac{1}{R_2}$. You need $R = 4.0\ \Omega$ and already have $R_1 = 12\ \Omega$. What must R_2 be?

Solution

Isolate the term containing the unknown, then combine over a common denominator:

$$\frac{1}{R_2} = \frac{1}{R} - \frac{1}{R_1} = \frac{R_1 - R}{R\,R_1}. \tag{C.1.3}$$

Only now flip both sides:

$$R_2 = \frac{R\,R_1}{R_1 - R} = \frac{(4.0)(12)}{12 - 4.0}\ \Omega = 6.0\,\Omega. \tag{C.1.4}$$

Check by substitution: $\frac{1}{12} + \frac{1}{6.0} = \frac{1}{12} + \frac{2}{12} = \frac{3}{12} = \frac{1}{4.0}$. ✓

Checkpoint C.1

The power delivered to a resistor is $P = I^2R$. Solve for the current I. How many solutions does the algebra give, and which do you keep if you only want the size of the current?

Answers to Checkpoints

Checkpoint C.1. Divide by R and take the square root: $I = \pm\sqrt{P/R}$. The algebra gives two solutions. The sign only tells the direction of the current, so for its size keep $I = \sqrt{P/R}$.

This page titled C.1: Solving an Equation for One Unknown is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.2: Quadratic Equations

An equation in which the unknown appears squared, and possibly also to the first power, is a **quadratic equation**. Quadratics arise constantly in kinematics, because position depends on t^2 under constant acceleration. Every quadratic can be rearranged into *standard form*, with everything on one side and zero on the other:

$$ax^2 + bx + c = 0. \tag{C.2.1}$$

(Here a, b and c are just the coefficients, not an acceleration or a speed of light.) A quadratic generally has *two* solutions, or *roots*. Which method is quickest depends on which coefficients are zero.

When one answer is zero

If there is no constant term ($c = 0$), every term contains x, and one root is $x = 0$. Do not divide by x; that would lose it. Instead, *factor out* x:

$$ax^2 + bx = 0 \quad \Longrightarrow \quad x\,(ax + b) = 0. \tag{C.2.2}$$

A product is zero only if one of its factors is zero, so either $x = 0$ or $ax + b = 0$:

$$x = 0 \qquad \text{or} \qquad x = -\frac{b}{a}. \tag{C.2.3}$$

Example C.3: How long is a kicked ball in the air?

A ball is kicked from level ground with an upward velocity component of $14.7\ \text{m/s}$. Its height is $y = (14.7\ \text{m/s})\ t - (4.9\ \text{m/s}^2)\ t^2$. When is it back on the ground?

Solution

The ball is on the ground when $y = 0$:

$$0 = 14.7\,t - 4.9\,t^2 = t\,(14.7 - 4.9\,t). \tag{C.2.4}$$

So $t = 0$ or $14.7 - 4.9\ t = 0$, which gives $t = 14.7/4.9 = 3.0\ \text{s}$. Both roots are physical: $t = 0$ is the moment of the kick, and $t = 3.0\ \text{s}$ is the landing. The zero root is not an error. It is the algebra correctly reporting that the ball also had $y = 0$ at the start.

A related shortcut: if there is no *first-power* term ($b = 0$), then $ax^2 + c = 0$ gives $x^2 = -c/a$ and $x = \pm\sqrt{-c/a}$, with no formula needed. A stone dropped from rest from a height of $20\ \text{m}$ lands when $0 = 20 - 4.9\ t^2$, so $t = \pm\sqrt{20/4.9} = \pm 2.0\ \text{s}$, and only the positive root is physical.

When neither answer is zero: the quadratic formula

When all three coefficients are nonzero, use the **quadratic formula**, which works for every quadratic:

Key Equation: The quadratic formula

$$ax^2 + bx + c = 0 \quad \Longrightarrow \quad x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{C.1}$$

The $\pm$ produces the two roots. The procedure, used in Example 2.4 of Chapter 2, is always the same:

1. Move every term to one side so the other side is zero.

2. Read off a, b and c *with their signs*. A wrong sign here is the most common error.
3. Substitute into equation (C.1) and compute both roots.
4. Decide which root answers the physical question. A negative time, for example, usually describes a moment before the motion began and is discarded.

Example C.4: A stone thrown from a cliff

A stone is thrown straight up at $12\ \mathrm{m/s}$ from the edge of a cliff $20\ \mathrm{m}$ above the beach. Taking the launch point as $y = 0$ and up as positive, its height is $y = 12\ t - 4.9\ t^2$ (in meters, with t in seconds). When does it hit the beach?

Solution

The beach is at $y = -20\ \mathrm{m}$:

$$-20 = 12\,t - 4.9\,t^2. \tag{C.2.5}$$

Move everything to the left side by adding $4.9\ t^2$ and subtracting $12\ t$:

$$4.9\,t^2 - 12\,t - 20 = 0, \qquad a = 4.9, \quad b = -12, \quad c = -20. \tag{C.2.6}$$

Then

$$t = \frac{-(-12) \pm \sqrt{(-12)^2 - 4(4.9)(-20)}}{2(4.9)} = \frac{12 \pm \sqrt{144 + 392}}{9.8} = \frac{12 \pm 23.15}{9.8}. \tag{C.2.7}$$

The roots are $t = 3.6\ \mathrm{s}$ and $t = -1.1\ \mathrm{s}$. The stone was thrown at $t = 0$, so the negative root is not part of this motion, and the stone lands at $t = 3.6\ \mathrm{s}$.

The negative root does mean something: it is when a stone on the same up-and-down path, coming up from the beach, would have had to leave the sand to be passing the cliff edge at $12\ \mathrm{m/s}$ at $t = 0$. The quadratic describes the whole path, and the physics decides which part of it is real.

When there is no answer

The quantity under the square root, $b^2 - 4ac$, is called the **discriminant**. If it is negative, the quadratic has no real roots, and in physics that is an answer in itself: the event never happens. Ask whether the stone in Example C.4 ever rises $10\ \mathrm{m}$ above the cliff edge. Setting $10 = 12\ t - 4.9\ t^2$ gives $4.9\ t^2 - 12\ t + 10 = 0$, with discriminant $144 - 196 = -52$. It is negative, so the stone never gets that high. (Its peak is $(12\,\mathrm{m/s})^2/(2 \times 9.8\ \mathrm{m/s^2}) = 7.3\ \mathrm{m}$ above the edge.)

Checkpoint C.2

Solve $3t^2 - 12t = 0$. A classmate divides both sides by t and gets only one answer. Which answer did they lose, and why?

Answers to Checkpoints

Checkpoint C.2. Factor: $3t(t - 4) = 0$, so $t = 0$ or $t = 4$. Dividing by t loses $t = 0$, because dividing by t quietly assumes $t \neq 0$.

This page titled C.2: Quadratic Equations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.3: Solving for Two Unknowns

Many problems give two unknowns at once: the tensions in two ropes, or the acceleration of a system and the tension in its string. One equation cannot pin down two unknowns, but two independent equations can. The rule to remember: *to solve for two unknowns, you need two equations*, and each one must give information the other does not. Two equations to be solved together are called **simultaneous equations**, and there are two standard ways to solve them. We show both on the pair

$$2x + 3y = 12, \qquad x - y = 1. \tag{C.3.1}$$

Substitution. Solve one equation for one unknown, and substitute the result into the other. The second equation gives $x = 1 + y$. Substituting into the first,

$$2(1 + y) + 3y = 12 \implies 2 + 5y = 12 \implies y = 2, \tag{C.3.2}$$

and then $x = 1 + y = 3$.

Elimination. Multiply one or both equations by constants so that adding or subtracting them cancels one unknown. Multiply the second equation by 3 to get $3x - 3y = 3$, and add it to the first:

$$(2x + 3y) + (3x - 3y) = 12 + 3 \implies 5x = 15 \implies x = 3, \tag{C.3.3}$$

and either equation then gives $y = 2$. Elimination is the method of choice when an unknown already appears with equal and opposite coefficients, as the tension does in the Atwood machine of Example 5.6.

Either way, finish by substituting the answer into *both* original equations: $2(3) + 3(2) = 12$ and $3 - 2 = 1$. ✓

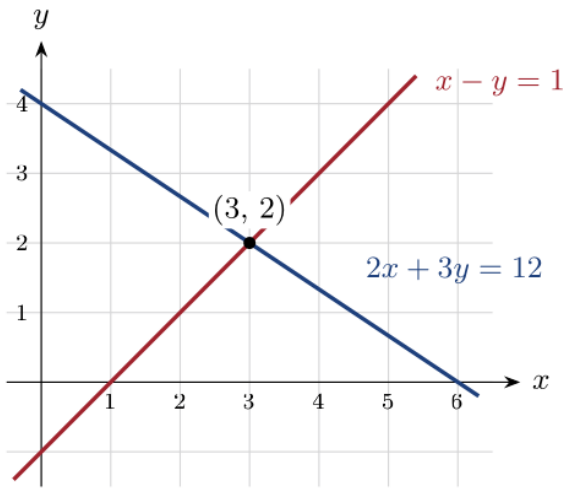


Figure C.1. Solving for two unknowns with two equations, seen as a picture. Each equation is a line of all the (x, y) pairs that satisfy it; the one pair that satisfies both is the point where the lines cross. Two parallel lines never cross (no solution), and two copies of the same line overlap everywhere (no unique solution).

•Figure description

A graph with x and y axes showing two straight lines. One line slopes downward from the point zero, four on the y-axis to the point six, zero on the x-axis; it is the equation two x plus three y equals twelve. The other slopes upward from zero, negative one through five, four; it is the equation x minus y equals one. The lines cross at a single marked point, three, two, which is the solution of the pair of equations.

Figure C.1 shows why two equations are needed and why they must be *independent*. If the second equation is just a multiple of the first, as $x + y = 10$ and $2x + 2y = 20$ are, it carries no new information, and the lines coincide. In physics this means you have written the same law twice and need another piece of physics, such as another direction of Newton's second law or another loop of a circuit, to close the problem.

Example C.5: A weight held by two ropes

A $50\ \mathrm{N}$ weight hangs from two ropes. Rope 1 pulls up and to the left at 30° above the horizontal; rope 2 pulls up and to the right at 60° above the horizontal. Find the tensions T_1 and T_2.

Solution

The weight is at rest, so the forces balance in each direction. Using the components of Section C.6, the horizontal parts of the tensions must cancel, and their vertical parts must hold up the weight:

$$\begin{aligned} x\text{:}&\quad T_2\cos 60^\circ - T_1\cos 30^\circ = 0,\\ y\text{:}&\quad T_1\sin 30^\circ + T_2\sin 60^\circ = 50\,\mathrm{N}. \end{aligned} \tag{C.3.4}$$

Two equations, two unknowns. Use substitution: the x-equation gives

$$T_1 = T_2\,\frac{\cos 60^\circ}{\cos 30^\circ} = 0.577\,T_2. \tag{C.3.5}$$

Substitute into the y-equation:

$$(0.577\,T_2)(0.500) + T_2\,(0.866) = 50 \implies 1.155\,T_2 = 50 \implies T_2 = 43.3\,\mathrm{N}, \tag{C.3.6}$$

and $T_1 = 0.577\ T_2 = 25.0\ \mathrm{N}$.

Check: horizontally, $43.3\cos 60^\circ = 21.7$ and $25.0\cos 30^\circ = 21.7$; vertically, $25.0\sin 30^\circ + 43.3\sin 60^\circ = 12.5 + 37.5 = 50$. ✓ The steeper rope carries more of the load, which makes sense. Chapter 5 solves a three-cable version of this problem in Example 5.5.

Checkpoint C.3

Why can't you solve the pair $x + y = 10$ and $2x + 2y = 20$ for a unique x and y? What does the graph of the two equations look like?

Answers to Checkpoints

Checkpoint C.3. The second equation is just twice the first, so it adds no new information: there is really only one equation for two unknowns. Graphically the two lines lie on top of each other, so every point on the line is a solution.

This page titled C.3: Solving for Two Unknowns is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.4: Solving for Three Unknowns

Circuits with more than one loop (Chapter 21) and some force problems have three unknowns. The rule is the same as before: *to solve for three unknowns, you need three independent equations*, one for each unknown. If you have only two, keep looking for another relation (another loop, another force equation, a constraint linking the motions) before you start solving. The method is the one you already know, applied one step further:

1. Use one equation to express one unknown in terms of the other two (or use elimination to remove it).
2. Substitute into the remaining two equations. You now have *two* equations for *two* unknowns.
3. Solve that pair by substitution or elimination.
4. *Back-substitute* the two values into the expression from step 1 to get the third unknown.
5. Check all three values in all three original equations.

Choose the unknown to remove wisely: pick the equation with the fewest terms, or an unknown whose coefficient is 1.

Example C.6: Currents in a two-loop circuit

Applying Kirchhoff's rules to a two-loop circuit gives the following equations for the three branch currents (in amperes; the numbers come from the batteries and resistors, as in Example 21.3):

$$\begin{aligned} &(1) \quad I_1 + I_2 = I_3, \\ &(2) \quad 2I_1 + 2I_3 = 8, \\ &(3) \quad 4I_2 + 2I_3 = 14. \end{aligned} \tag{C.4.1}$$

Find I_1, I_2 and I_3.

Solution

Steps 1–2: reduce to two unknowns. Equations (2) and (3) each contain only two currents, so solve each for the current that is not I_3:

$$\text{from (2):} \quad I_1 = 4 - I_3, \qquad \text{from (3):} \quad I_2 = 3.5 - 0.5\,I_3. \tag{C.4.2}$$

Substitute both into equation (1), which leaves a single unknown:

$$(4 - I_3) + (3.5 - 0.5\,I_3) = I_3 \implies 7.5 = 2.5\,I_3 \implies I_3 = 3.0\,\text{A}. \tag{C.4.3}$$

(Here the reduction went straight from three unknowns to one, because two of the equations each had only two unknowns. In general you will pass through a two-by-two system on the way.)

Step 4: back-substitute. $I_1 = 4 - 3.0 = 1.0\ \text{A}$ and $I_2 = 3.5 - 0.5(3.0) = 2.0\ \text{A}$.

Step 5: check. (1) $1.0 + 2.0 = 3.0$; (2) $2(1.0) + 2(3.0) = 8$; (3) $4(2.0) + 2(3.0) = 14$. ✓

For larger systems, or for messy numbers, the same steps can be organized as a matrix and handed to a calculator or computer. The physics is in writing down the equations; the solving is bookkeeping. If a current comes out negative, that is not an error either: it means the current flows opposite to the direction you assumed.

This page titled C.4: Solving for Three Unknowns is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.5: Right-Triangle Trigonometry

A **right triangle** has one 90° angle. The side opposite the right angle is the longest side, the **hypotenuse**. Pick one of the other two angles and call it θ. Relative to θ, the shorter sides are named for where they sit: the *opposite* side faces θ, and the *adjacent* side touches it (Figure C.2).

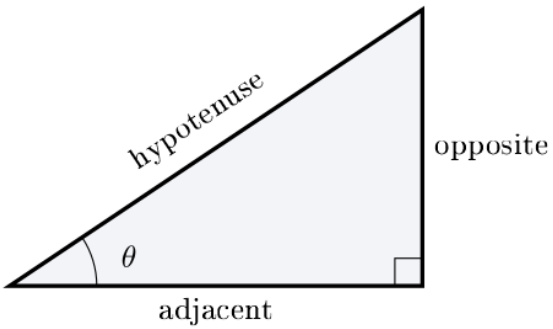


Figure C.2. The sides of a right triangle, named relative to the angle θ. The hypotenuse is always opposite the right angle; which leg is "opposite" and which is "adjacent" depends on which angle you call θ.

•Figure description

A right triangle with its right angle at the lower right corner, marked with a small square. The angle theta is at the lower left corner. The horizontal bottom side, touching theta, is labeled adjacent. The vertical right side, across from theta, is labeled opposite. The long slanted side from the lower left to the upper right is labeled hypotenuse.

The three **trigonometric functions** are ratios of these sides:

Key Equation: Right-triangle trigonometry

$$\sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}, \qquad \cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}}, \qquad \tan\theta = \frac{\text{opposite}}{\text{adjacent}}. \tag{C.2}$$

Many students remember these as "SOH CAH TOA": **S**ine is **O**pposite over **H**ypotenuse, and so on. The sides are also linked by the **Pythagorean theorem**:

Key Equation: Pythagorean theorem

$$(\text{hypotenuse})^2 = (\text{opposite})^2 + (\text{adjacent})^2. \tag{C.3}$$

Together these let you *solve* a right triangle, finding every side and angle, from any two pieces of information, as long as at least one of them is a side:

- **Two sides known:** the Pythagorean theorem gives the third side, and an inverse trig function gives the angle.
- **One side and one angle known:** rearrange equation (C.2) to get each of the other sides.

To find an angle from a ratio, use an **inverse trigonometric function** ($\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$, also written arcsin, arccos, arctan). For example, if the opposite side is 3 and the adjacent side is 4, then $\theta = \tan^{-1}(3/4) = 36.9^\circ$. Make sure your calculator is in *degree* mode when you want degrees.

The two non-right angles of a right triangle add up to 90°, so if one is θ the other is $90^\circ - \theta$. Viewed from that other angle, "opposite" and "adjacent" trade places, which is why $\sin\theta = \cos(90^\circ - \theta)$. It is also the source of the most common mistake in this whole appendix: using $\sin$ where $\cos$ belongs because the angle was measured from a different side than you assumed. Always ask which side the angle actually touches.

A few values are worth knowing by heart, both for exact work and for quick estimates:

Sine, cosine and tangent of common angles

θ	0°	30°	45°	60°	90°
$\sin\theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2} \approx 0.707$	$\frac{\sqrt{3}}{2} \approx 0.866$	1
$\cos\theta$	1	$\frac{\sqrt{3}}{2} \approx 0.866$	$\frac{\sqrt{2}}{2} \approx 0.707$	$\frac{1}{2}$	0
$\tan\theta$	0	$\frac{1}{\sqrt{3}} \approx 0.577$	1	$\sqrt{3} \approx 1.73$	undefined

Example C.7: How long is the ramp?

A wheelchair ramp must rise $1.2\ \text{m}$ at an angle of 5.0° above the horizontal. How long is the ramp surface, and how much horizontal floor space does it take up?

Solution

The ramp surface is the hypotenuse. The rise is the side opposite the 5.0° angle, and the floor space is the adjacent side. Sine connects the opposite side to the hypotenuse:

$$\sin\theta = \frac{\text{rise}}{L} \implies L = \frac{\text{rise}}{\sin\theta} = \frac{1.2\,\text{m}}{\sin 5.0^\circ} = \frac{1.2\,\text{m}}{0.0872} = 13.8\,\text{m}. \tag{C.5.1}$$

Tangent connects the opposite and adjacent sides:

$$\text{run} = \frac{\text{rise}}{\tan\theta} = \frac{1.2\,\text{m}}{0.0875} = 13.7\,\text{m}. \tag{C.5.2}$$

Check with the Pythagorean theorem: $\sqrt{(13.7)^2 + (1.2)^2} = \sqrt{187.7 + 1.44} = 13.8\ \text{m}$. ✓ For such a shallow angle the hypotenuse is only slightly longer than the run, as you would expect.

This page titled C.5: Right-Triangle Trigonometry is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.6: Breaking a Vector into Components

Right-triangle trigonometry earns its keep in physics through *vectors*. Draw a vector $\mathbf{A}$, then drop a line from its tip straight down to the x-axis. You have made a right triangle whose hypotenuse is the vector itself and whose legs are its *components* A_x and A_y (Figure C.3). If θ is measured from the $+x$ axis, A_x is the adjacent side and A_y the opposite side, so

$$A_x = A\cos\theta, \qquad A_y = A\sin\theta, \tag{C.6.1}$$

which are equations (3.1) and (3.2) of Chapter 3. Breaking a vector into components replaces one slanted arrow by two perpendicular ones that add up to it. Their great advantage is that x-parts and y-parts can be handled separately, by ordinary arithmetic.

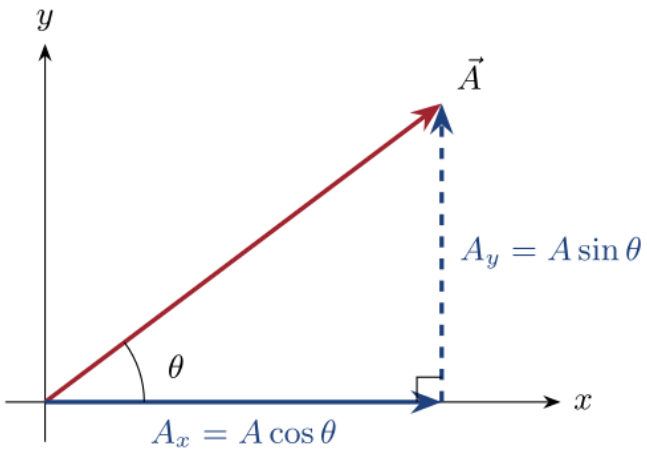


Figure C.3. A vector and its components form a right triangle. The vector is the hypotenuse; with θ measured from the $+x$ axis, the x -component is the adjacent side and the y-component is the opposite side.

•Figure description

A vector A drawn as an arrow from the origin up and to the right, at angle theta above the positive x-axis. A blue horizontal arrow along the x-axis from the origin to directly below the tip is labeled A sub x equals A cosine theta. A dashed vertical arrow from the end of A sub x up to the tip of A is labeled A sub y equals A sine theta. The vector and its two components form a right triangle with the vector as hypotenuse.

Example C.8: Pulling a sled

A rope pulls a sled with a force of $25\ \mathrm{N}$ directed 40° above the horizontal. Find the horizontal and vertical components of the force.

Solution

The angle is measured from the horizontal, so the horizontal component is adjacent to it:

$$F_x = F\cos\theta = (25\,\mathrm{N})\cos 40^\circ = 19\,\mathrm{N}, \qquad F_y = F\sin\theta = (25\,\mathrm{N})\sin 40^\circ = 16\,\mathrm{N}. \tag{C.6.2}$$

Check: each component is smaller than $25\ \mathrm{N}$, as legs of a right triangle must be, and $\sqrt{19.15^2 + 16.07^2} = 25.0\ \mathrm{N}$. ✓

When the angle is not measured from the x-axis

The formulas $A\cos\theta$ and $A\sin\theta$ are not rules about x and y; they are rules about *adjacent* and *opposite*. The component next to the angle always gets the cosine. If an angle is measured from the vertical, the y-component gets the cosine and the x-component the sine.

The most important case is an object on an incline. It is convenient to use axes along the slope and perpendicular to it, and then the weight mg must be broken into components along those tilted axes. The angle between the weight and the perpendicular axis equals the incline angle θ, because both angles are what is left over when the same angle is subtracted from 90° (Figure C.4). The perpendicular component touches that angle and gets the cosine:

$$\text{along the slope (downhill):} \quad mg\sin\theta, \qquad \text{perpendicular, into the surface:} \quad mg\cos\theta. \tag{C.6.3}$$

Check with limits: on flat ground ($\theta = 0$) the weight pushes entirely into the surface, $mg\cos 0 = mg$, with no downhill pull. On a vertical wall ($\theta = 90^\circ$) the weight is entirely along the "slope," $mg\sin 90^\circ = mg$. If you ever swap sine and cosine on an incline, this check will catch it.

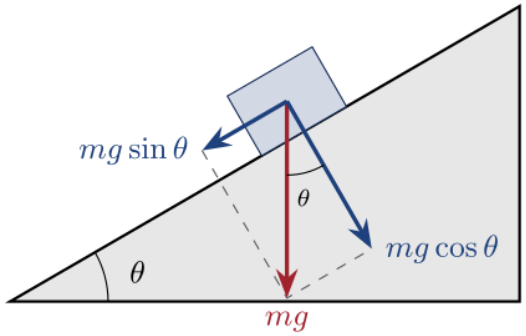


Figure C.4. Components of the weight on an incline of angle θ. The weight mg is the hypotenuse. The angle between the weight and the perpendicular to the surface equals θ, so the perpendicular component (adjacent to θ) is $mg\cos\theta$ and the component down the slope (opposite θ) is $mg\sin\theta$.

•Figure description

A ramp rising to the right at angle theta above the horizontal ground, with a block resting on it. From the center of the block, a red arrow points straight down, labeled m g. Two blue arrows from the same point show its components: one points down the slope, labeled m g sine theta, and one points perpendicular into the ramp surface, labeled m g cosine theta. Dashed lines complete the right triangle, with the weight as its hypotenuse. A small arc marks the angle theta between the weight and the perpendicular component, equal to the ramp angle, which is also marked at the bottom of the ramp.

Checkpoint C.4

A $10\ \text{m}$ displacement points 30° to the right of the $+y$ axis (that is, 30° east of north). Find its x- and y-components. Which one gets the cosine, and why?

Answers to Checkpoints

Checkpoint C.4. The angle is measured from the y-axis, so the y-component is adjacent to it and gets the cosine: $A_y = (10\ \text{m})\cos 30^\circ = 8.7\ \text{m}$ and $A_x = (10\ \text{m})\sin 30^\circ = 5.0\ \text{m}$.

This page titled C.6: Breaking a Vector into Components is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.7: Finding the Hypotenuse from the Components

Going the other way, from components back to the vector, is solving a right triangle whose two legs are known. The magnitude is the hypotenuse, from the Pythagorean theorem, and the direction comes from the tangent:

Key Equation: Magnitude and direction from components

$$A = \sqrt{A_x^2 + A_y^2}, \qquad \theta = \tan^{-1}\left(\frac{A_y}{A_x}\right). \tag{C.4}$$

These are equations (3.3) and (3.4) of Chapter 3. One warning about the angle. A calculator's $\tan^{-1}$ always returns an angle between -90° and $+90^\circ$, that is, a direction pointing to the right. If A_x is negative, the vector points to the left, and you must add 180° to the calculator's answer. A quick sketch showing which quadrant the vector lies in (Figure 3.4) prevents this error.

Example C.9: Where did the hiker end up?

(a) A hiker walks $3.0\ \text{km}$ east and then $4.0\ \text{km}$ north. How far is she from her starting point, and in what direction? (b) A puck slides with velocity components $v_x = -6.0\ \text{m/s}$ and $v_y = +8.0\ \text{m/s}$. Find its speed and direction.

Solution

(a) The east and north legs are the components, and the straight-line distance is the hypotenuse:

$$d = \sqrt{(3.0)^2 + (4.0)^2}\ \text{km} = \sqrt{25}\ \text{km} = 5.0\,\text{km}, \tag{C.7.1}$$

in the direction

$$\theta = \tan^{-1}\left(\frac{4.0}{3.0}\right) = 53.1^\circ\ \text{north of east}. \tag{C.7.2}$$

(b) The speed is the magnitude, $v = \sqrt{(-6.0)^2 + (8.0)^2} = 10.0\ \text{m/s}$. The calculator gives $\tan^{-1}\big(8.0/(-6.0)\big) = -53.1^\circ$, which points down and to the right. But the puck moves up and to the left ($v_x < 0$, $v_y > 0$), so add 180°: $\theta = 126.9^\circ$, measured counterclockwise from the $+x$ axis. That is, 53.1° above the $-x$ axis.

If you know one leg and the angle rather than both legs, use trigonometry for the hypotenuse instead, as in Example C.7: $\text{hypotenuse} = \text{opposite}/\sin\theta = \text{adjacent}/\cos\theta$. Either route must give the same answer, which makes a handy check.

This page titled C.7: Finding the Hypotenuse from the Components is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.8: Sinusoids and Phasors

Anything that oscillates smoothly, such as the charge on a capacitor in an LC circuit, the voltage at a wall outlet, or the displacement of a string carrying a wave, is described by a sine or a cosine:

$$y = A\sin(\omega t + \phi). \tag{C.8.1}$$

Three numbers pin it down. The *amplitude* A is the largest value y reaches. The *angular frequency* ω, in radians per second, says how quickly the cycle repeats; it is related to the frequency f and the period T by $\omega = 2\pi f = 2\pi/T$. The *phase constant* ϕ says where in its cycle the oscillation is at $t = 0$. The whole argument $\omega t + \phi$ is called the *phase*.

A rotating vector

Picture a vector of length A with its tail at the origin, turning counterclockwise at a steady angular speed ω, so that at time t it makes the angle $\omega t + \phi$ with the $+x$ axis (Figure C.5). By Section C.6, its y-component is

$$A\sin(\omega t + \phi), \tag{C.8.2}$$

which is exactly our sinusoid. As the vector turns, its y-component rises to A when the vector points straight up, falls through zero when it points to the left, reaches $-A$ when it points straight down, and returns to zero as the vector completes the turn. One revolution is one cycle. A vector used this way is called a **phasor**.

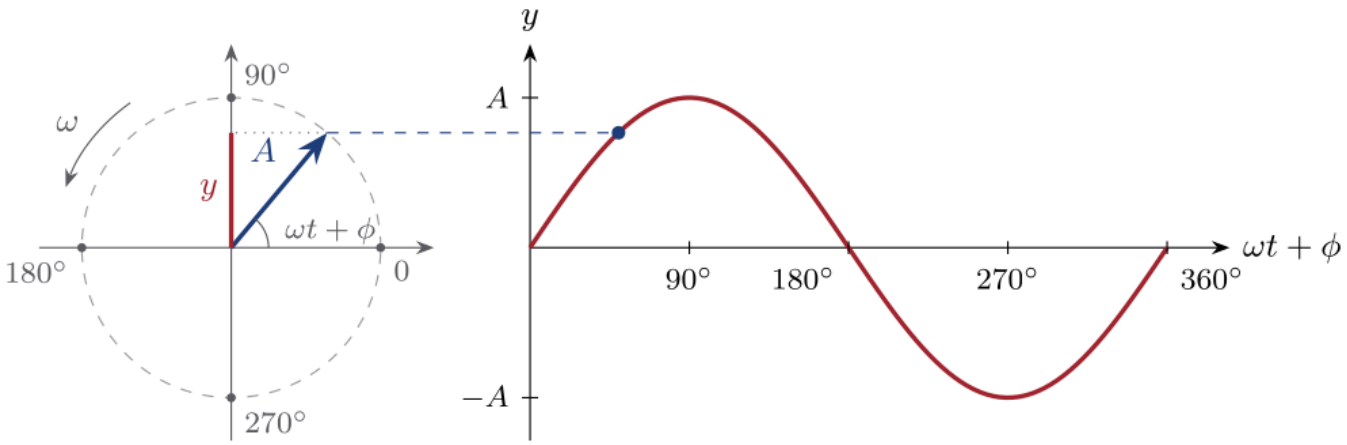


Figure C.5. A phasor is a vector of length A that rotates counterclockwise with angular speed ω. Its vertical component $y = A\sin(\omega t + \phi)$ traces out a sine curve, one full cycle for each revolution.

•Figure description

Left: a phasor diagram. A blue arrow of length A starts at the origin and points up and to the right, at the angle omega t plus phi above the positive x-axis. A dashed circle of radius A shows the path of its tip, with dots marking zero, 90, 180, and 270 degrees, and a curved arrow labeled omega shows that it rotates counterclockwise. A red segment along the vertical axis, from the origin up to the height of the arrow's tip, is labeled y, the arrow's vertical component. Right: a graph of y against the angle omega t plus phi from 0 to 360 degrees, a red sine curve between A and minus A. A dashed horizontal line runs from the tip of the phasor across to a dot on the curve at the same height, showing that the curve is traced out by the phasor's vertical component as it turns.

Note: The phasor is not the quantity

A phasor is a bookkeeping device, not a physical arrow in space. Only its vertical component is the actual voltage, current, or displacement at that instant. The advantage of the picture is that it turns statements about *timing*, which are hard to see in formulas, into statements about *angles*, which are easy to see in a drawing.

Phase differences

Two sinusoids with the same angular frequency,

$$y_1 = A_1 \sin \omega t, \qquad y_2 = A_2 \sin(\omega t + \phi), \tag{C.8.3}$$

are drawn as two phasors that turn together, always separated by the same angle ϕ (Figure C.6). The whole diagram rotates rigidly, so the angle between the phasors never changes. That fixed angle is the *phase difference*. The phasor that is ahead in the counterclockwise direction reaches the top of the circle first, so its sinusoid reaches its peak first; we say it *leads*, and the other *lags*. In time, a phase difference ϕ (in radians) is a delay of

$$\Delta t = \frac{\phi}{\omega} = \frac{\phi}{2\pi} T, \tag{C.8.4}$$

so a phase difference of 90° $(\pi/2)$ is a quarter period. Since $\cos \omega t = \sin(\omega t + 90^\circ)$, a cosine is just a sine that leads by a quarter cycle.

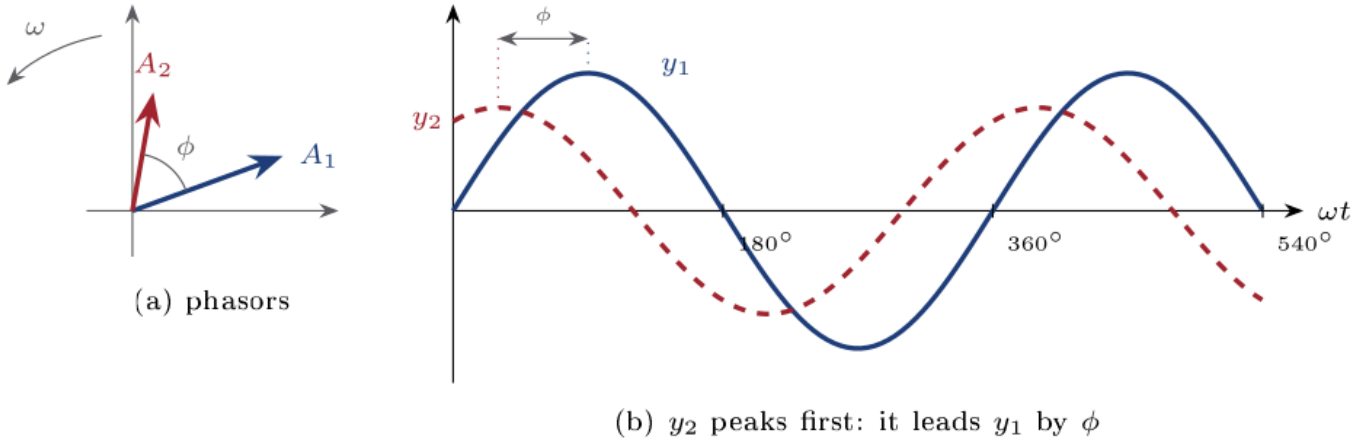


Figure C.6. Two sinusoids of the same frequency differing in phase by ϕ. (a) Their phasors rotate together with a fixed angle ϕ between them. (b) The phasor that is ahead counterclockwise belongs to the sinusoid that peaks first.

•Figure description

Panel a: two phasors from the origin. A blue phasor of length A sub 1 points a little above the positive x-axis. A shorter red phasor of length A sub 2 points nearly straight up, ahead of the blue one in the counterclockwise direction by the angle phi, which is marked with an arc. A curved arrow labeled omega shows that both rotate counterclockwise together. Panel b: graphs against omega t of y sub 1, a solid blue sine curve, and y sub 2, a dashed red sine curve of smaller amplitude shifted to the left. Dotted lines mark the first peak of each curve, and a double arrow labeled phi between them shows that y sub 2 reaches its peak earlier: y sub 2 leads y sub 1 by phi.

Adding sinusoids

The real power of phasors appears when sinusoids must be *added*, as happens with the voltages around an ac circuit or with two waves arriving at the same point. The sum $y_1 + y_2$ is the sum of the y-components of the two phasors. But the y-component of a vector sum is the sum of the y-components, so $y_1 + y_2$ is the y-component of the *vector sum* of the two phasors. That vector sum is itself a phasor: it has a fixed length and turns rigidly with the other two. So the sum of two sinusoids of the same frequency is another sinusoid of that frequency, and its amplitude and phase come from ordinary vector addition, done at any convenient instant, usually $t = 0$.

Put A_1 along the x-axis and A_2 at the angle ϕ. The components of the sum are $A_1 + A_2 \cos \phi$ and $A_2 \sin \phi$, and Section C.7 gives its length and angle:

Key Equation: Adding two sinusoids of the same frequency

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi}, \qquad \tan\theta = \frac{A_2\sin\phi}{A_1 + A_2\cos\phi}, \tag{C.5}$$

so that $y_1 + y_2 = A\sin(\omega t + \theta)$. The amplitudes add directly only when $\phi = 0$; in general the peaks of y_1 and y_2 happen at different times, and the sum's peak is smaller than $A_1 + A_2$. Three special cases are worth remembering:

- $\phi = 0$ (in phase): $A = A_1 + A_2$.
- $\phi = 180^\circ$ (opposite phase): $A = |A_1 - A_2|$, which is zero if the amplitudes are equal.
- $A_1 = A_2$: $A = 2A_1\cos(\phi/2)$, the result used for waves in Chapter 29 and for light in Chapter 33.

Phasors are used this way for ac circuits in Chapter 26, for adding waves in Chapter 29, and for interference and diffraction in Chapters 33 and 34.

Example C.10: Adding two ac voltages

Two circuit elements in series have voltages $v_1 = (3.0\ \text{V})\sin\omega t$ and $v_2 = (4.0\ \text{V})\sin(\omega t + 90^\circ)$. Find the total voltage $v_1 + v_2$.

Solution

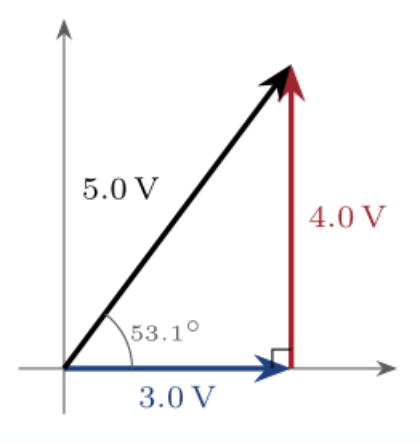
The phasors are perpendicular. Drawn head to tail, they form a right triangle (below), so the amplitude is the hypotenuse,

$$V = \sqrt{(3.0)^2 + (4.0)^2}\ \text{V} = 5.0\,\text{V}, \tag{C.8.5}$$

and the angle is $\theta = \tan^{-1}(4.0/3.0) = 53.1^\circ$. Equation (C.5) with $\phi = 90^\circ$ gives the same thing, since $\cos 90^\circ = 0$. So

$$v_1 + v_2 = (5.0\,\text{V})\sin(\omega t + 53.1^\circ). \tag{C.8.6}$$

Check: at $t = 0$, $v_1 = 0$ and $v_2 = 4.0$ V, and the answer gives $(5.0)\sin 53.1^\circ = 4.0$ V. ✓ Notice that the total amplitude is 5.0 V, not 7.0 V: the two voltages never reach their peaks at the same moment.



•Figure description
Phasor addition of two perpendicular phasors. A blue arrow of length 3.0 volts points to the right along the x-axis from the origin. From its tip, a red arrow of length 4.0 volts points straight up, with a small square marking the right angle. A black arrow from the origin to the tip of the red arrow is the sum, 5.0 volts long, at 53.1 degrees above the x-axis.

Checkpoint C.5

Two voltages of the same frequency have amplitudes 6.0 V and 8.0 V. (a) What are the largest and smallest possible amplitudes of their sum, and what phase differences give them? (b) What phase difference makes the sum's amplitude 10.0 V?

Answers to Checkpoints

Checkpoint C.5. (a) The largest is $6.0 + 8.0 = 14.0\ \text{V}$, when they are in phase ($\phi = 0$). The smallest is $8.0 - 6.0 = 2.0\ \text{V}$, when they are opposite ($\phi = 180^\circ$). (b) $10.0^2 = 6.0^2 + 8.0^2 + 2(6.0)(8.0)\cos\phi$ gives $100 = 100 + 96\cos\phi$, so $\cos\phi = 0$ and $\phi = 90^\circ$. The phasors form a 6–8–10 right triangle.

This page titled C.8: Sinusoids and Phasors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.S: Basic Algebra and Trigonometry for Physics Students (Summary)

Summary

- To solve for an unknown, do the same operation to both sides, undoing what was done to the unknown in reverse order. Solve with symbols first; then check units and substitute the answer back.
- A quadratic $ax^2 + bx + c = 0$ has two roots. If $c = 0$, factor out x: the roots are 0 and $-b/a$ (never divide by x). Otherwise use the quadratic formula, reading off a, b, c with their signs, and let the physics choose the root. A negative discriminant means the event never happens.
- Two unknowns need two independent equations. Solve by substitution (solve one equation for one unknown and substitute) or elimination (add or subtract multiples of the equations to cancel an unknown). Graphically, the solution is where two lines cross.
- Three unknowns need three independent equations. Remove one unknown to get a two-by-two system, solve it, back-substitute, and check all three equations.
- In a right triangle, $\sin\theta = \text{opp}/\text{hyp}$, $\cos\theta = \text{adj}/\text{hyp}$, $\tan\theta = \text{opp}/\text{adj}$, and $\text{hyp}^2 = \text{opp}^2 + \text{adj}^2$.
- A vector is the hypotenuse of a triangle whose legs are its components. The component adjacent to the angle gets the cosine. On an incline, the weight's components are $mg\sin\theta$ along the slope and $mg\cos\theta$ into the surface.
- From components, $A = \sqrt{A_x^2 + A_y^2}$ and $\theta = \tan^{-1}(A_y/A_x)$, adding 180° when $A_x < 0$.
- A sinusoid $A\sin(\omega t + \phi)$ is the vertical component of a phasor, a vector of length A rotating counterclockwise at angular speed ω. The phasor ahead counterclockwise leads; a phase difference ϕ is a time shift ϕ/ω.
- Sinusoids of the same frequency add like their phasors, as vectors: $A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi}$. Amplitudes add directly only when the sinusoids are in phase.

Key Equations

Quadratic with one zero root

$$ax^2 + bx = 0 \implies x = 0 \quad \text{or} \quad x = -\frac{b}{a} \tag{C.S.1}$$

The quadratic formula

$$ax^2 + bx + c = 0 \implies x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{C.S.2}$$

Right-triangle trigonometry

$$\sin\theta = \frac{\text{opp}}{\text{hyp}}, \qquad \cos\theta = \frac{\text{adj}}{\text{hyp}}, \qquad \tan\theta = \frac{\text{opp}}{\text{adj}} \tag{C.S.3}$$

Pythagorean theorem

$$\text{hyp}^2 = \text{opp}^2 + \text{adj}^2 \tag{C.S.4}$$

Components, and back

$$A_x = A\cos\theta, \quad A_y = A\sin\theta; \qquad A = \sqrt{A_x^2 + A_y^2}, \quad \theta = \tan^{-1}\frac{A_y}{A_x} \tag{C.S.5}$$

Adding two sinusoids

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\phi}, \qquad \tan\theta = \frac{A_2\sin\phi}{A_1 + A_2\cos\phi} \tag{C.S.6}$$

This page titled C.S: Basic Algebra and Trigonometry for Physics Students (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

C.E: Basic Algebra and Trigonometry for Physics Students (Practice Problems)

Answers follow the problems.

Solving for One Unknown

C.1 ★ Solve the kinetic energy equation $K = \frac{1}{2}mv^2$ for v. Then find v for $K = 200\ \text{J}$ and $m = 4.0\ \text{kg}$.

C.2 ★ The period of a pendulum is $T = 2\pi\sqrt{L/g}$. Solve for the length L, and find the length of a pendulum with a period of $2.0\ \text{s}$ ($g = 9.80\ \text{m/s}^2$).

C.3 ★★ The thin-lens equation is $\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}$. Solve for d_i, then evaluate it for $f = 10\ \text{cm}$ and $d_o = 15\ \text{cm}$. Check your answer by substituting it back.

Quadratic Equations

C.4 ★ Solve $5t^2 - 20t = 0$. Interpret both roots if t is the time at which a thrown object is at its launch height.

C.5 ★ Solve $2x^2 - 3x - 5 = 0$ with the quadratic formula, and check both roots by substitution.

C.6 ★★ A ball leaves a child's hand $2.0\ \text{m}$ above the ground, moving straight up at $15\ \text{m/s}$. Its height is $y = 2.0 + 15\ t - 4.9\ t^2$ (meters, seconds). When does it hit the ground?

C.7 ★★ Does the ball of the previous problem ever reach $15\ \text{m}$ above the ground? Answer using the discriminant, and then find the greatest height it does reach.

Two and Three Unknowns

C.8 ★ Solve the pair $x + 2y = 7$ and $3x - y = 7$, once by substitution and once by elimination.

C.9 ★★ A $100\ \text{N}$ sign hangs from two cables. Cable 1 pulls up and to the left at 45° above horizontal; cable 2 pulls up and to the right at 30° above horizontal. Write the x- and y-equations for equilibrium and solve for both tensions.

C.10 ★★★ Solve the system $x + y + z = 6$, $2x - y + z = 3$, $x + 2y - z = 2$.

C.11 ★★★ Kirchhoff's rules for a circuit give $I_1 = I_2 + I_3$, $2I_1 + 4I_2 = 14$, and $4I_2 - 8I_3 = 0$ (currents in amperes). Find all three currents.

Right Triangles and Components

C.12 ★ The hypotenuse of a right triangle is $13\ \text{m}$ and one leg is $5.0\ \text{m}$. Find the other leg and the smaller of the two acute angles.

C.13 ★ A guy wire runs from the ground to a point $8.0\ \text{m}$ up a vertical pole, making a 60° angle with the ground. How long is the wire? Find it with trigonometry, then check with the Pythagorean theorem.

C.14 ★ A $40\ \text{N}$ force points at 120° counterclockwise from the $+x$ axis. Find its x- and y-components, with signs.

C.15 ★★ An object moves with velocity components $v_x = -3.0\ \text{m/s}$ and $v_y = -4.0\ \text{m/s}$. Find its speed and direction, measured counterclockwise from the $+x$ axis.

C.16 ★★ A $5.0\ \text{kg}$ crate rests on a ramp inclined at 25°. Find the components of its weight along the ramp and perpendicular to it ($g = 9.80\ \text{m/s}^2$).

Sinusoids and Phasors

C.17 ★ A current is $i = (2.0\ \text{A})\sin(377t + 30^\circ)$, with t in seconds. Find its amplitude, its frequency, and its value at $t = 0$.

C.18 ★ The voltages $v_1 = V_m \sin\omega t$ and $v_2 = V_m \cos\omega t$ have a frequency of $50\ \text{Hz}$. Which one leads, and by how much time?

C.19 ★★ Add $v_1 = (12\ \text{V})\sin\omega t$ and $v_2 = (5.0\ \text{V})\sin(\omega t - 90^\circ)$. Give the amplitude of the sum and its phase relative to v_1. Does the sum lead or lag v_1?

C.20 ★★ Two waves of equal amplitude A arrive at a point with a phase difference of 120°. Find the amplitude of their sum, once with the general formula and once with $2A\cos(\phi/2)$.

Answers to Practice Problems

C.1 $v = \sqrt{2K/m} = \sqrt{2(200)/4.0} = 10\ \text{m/s}$.

C.2 $L = gT^2/(4\pi^2) = (9.80)(2.0)^2/(4\pi^2) = 0.99\ \text{m}$.

C.3 $d_i = \dfrac{f\,d_o}{d_o - f} = \dfrac{(10)(15)}{15 - 10} = 30\ \text{cm}$. Check: $\frac{1}{15} + \frac{1}{30} = \frac{3}{30} = \frac{1}{10}$.

C.4 $5t(t-4) = 0$, so $t = 0$ (launch) and $t = 4$ (back at launch height).

C.5 $x = \dfrac{3 \pm \sqrt{9 + 40}}{4} = \dfrac{3 \pm 7}{4}$, so $x = 2.5$ or $x = -1$.

C.6 $4.9\ t^2 - 15\ t - 2.0 = 0$ gives $t = (15 \pm 16.25)/9.8$; the physical root is $t = 3.2\ \text{s}$ (the other is $-0.13\ \text{s}$).

C.7 $4.9\ t^2 - 15\ t + 13 = 0$ has discriminant $225 - 254.8 = -29.8 < 0$, so no. The greatest height is $2.0 + (15)^2/(2 \times 9.8) = 13.5\ \text{m}$.

C.8 $x = 3$, $y = 2$.

C.9 $T_1\cos 45^\circ = T_2 \cos 30^\circ$ and $T_1 \sin 45^\circ + T_2 \sin 30^\circ = 100$; $T_1 = 89.7\ \text{N}$, $T_2 = 73.2\ \text{N}$.

C.10 $x = 1$, $y = 2$, $z = 3$.

C.11 $I_1 = 3.0\ \text{A}$, $I_2 = 2.0\ \text{A}$, $I_3 = 1.0\ \text{A}$.

C.12 The other leg is $\sqrt{13^2 - 5.0^2} = 12\ \text{m}$; the smaller angle is $\sin^{-1}(5.0/13) = 22.6^\circ$.

C.13 $L = 8.0/\sin 60^\circ = 9.2\ \text{m}$. As a check, the base is $8.0/\tan 60^\circ = 4.62\ \text{m}$, and $\sqrt{8.0^2 + 4.62^2} = 9.2\ \text{m}$.

C.14 $F_x = 40\cos 120^\circ = -20\ \text{N}$, $F_y = 40\sin 120^\circ = +34.6\ \text{N}$.

C.15 $v = 5.0\ \text{m/s}$. The calculator gives $\tan^{-1}(4.0/3.0) = 53.1^\circ$, but both components are negative, so $\theta = 180^\circ + 53.1^\circ = 233.1^\circ$.

C.16 $mg = 49.0\ \text{N}$; along the ramp $mg\sin 25^\circ = 20.7\ \text{N}$; perpendicular $mg\cos 25^\circ = 44.4\ \text{N}$.

C.17 The amplitude is $2.0\ \text{A}$; $f = 377/2\pi = 60\ \text{Hz}$; $i(0) = 2.0\sin 30^\circ = 1.0\ \text{A}$.

C.18 $\cos\omega t = \sin(\omega t + 90^\circ)$, so v_2 leads by a quarter period: $T/4 = (1/50)/4 = 5.0\ \text{ms}$.

C.19 $V = \sqrt{12^2 + 5.0^2} = 13\ \text{V}$, at $\theta = \tan^{-1}(-5.0/12) = -22.6^\circ$: the sum lags v_1 by 22.6°.

C.20 $\sqrt{A^2 + A^2 + 2A^2\cos 120^\circ} = \sqrt{A^2} = A$, and $2A\cos 60^\circ = A$. The sum has the same amplitude as either wave.

This page titled C.E: Basic Algebra and Trigonometry for Physics Students (Practice Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D: MOTIVATION OF CALCULUS FOR PHYSICS STUDENTS

Physics is about how things change. A car speeds up, a capacitor charges, a spring stretches, a planet swings around the Sun. Almost every time we describe such a change, one of two questions comes up:

1. *How fast is this quantity changing right now?*
2. *How much of this quantity has built up in total?*

The **calculus** is the mathematics of exactly these two questions. The first is answered by the *derivative*, the second by the *integral*. This appendix is not a calculus course. It does not teach many techniques, and you will not need many in this book. Its goal is that whenever you meet a d or a $\int$ in a physics equation, you know what it *means*. Two ideas carry nearly all of that meaning:

- A derivative is a **rate of change**. On a graph it is a *slope*.
- An integral is a **sum of many small pieces**. On a graph it is an *area*.

Learning Objectives

- Interpret a derivative as the slope of a graph and as a rate of change, and state its units.
- See how the slope at a single instant arises from average rates over shrinking intervals.
- Interpret an integral as the limit of a sum of small pieces, and as the area under a graph.
- Explain why differentiation and integration undo each other, and use that to evaluate simple integrals.
- Set up a physics integral by slicing an object into small pieces labeled by a coordinate.
- Check a calculus result with units, signs, limiting cases, and a rough average.

Note: A car already does calculus

Look at a car's dashboard. The *odometer* tells you how far the car has gone in total. The *speedometer* tells you how fast that odometer reading is changing at this moment. The speedometer reading is the derivative of the odometer reading; the odometer reading is the integral of the speedometer reading. Keep this picture in mind. Everything in this appendix is an elaboration of it.

Close message
Learn more
Welcome to your new guide!
Click on "New" to build out your guide.

D.1: The Derivative: Slope as a Rate of Change
D.2: The Integral: Adding Up Small Pieces
D.3: Why the Two Ideas Undo Each Other
D.4: Integration in Physics: Slice, Label, Add
D.5: Checking a Calculus Result
D.S: Motivation of Calculus for Physics Students (Summary)
D.E: Motivation of Calculus for Physics Students (Practice Problems)

This page titled D: Motivation of Calculus for Physics Students is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.1: The Derivative: Slope as a Rate of Change

Average rate: rise over run

Suppose a cart rolls along a track, and its position x is recorded at every time t. If it is at x_1 at time t_1 and at x_2 at time t_2, its average velocity over that interval is

$$v_{\text{avg}} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}. \tag{D.1.1}$$

This is the change in one quantity divided by the change in another: a **rate of change**. On a graph of x versus t it is *rise over run*, the **slope** of the straight line joining the two points. A line that cuts a curve at two points like this is called a **secant line**.

If the graph is itself a straight line, that is the end of the story: the slope is the same everywhere and so is the velocity. But most graphs curve. For a curved graph the average depends on which interval you choose, and it blurs together everything that happened inside it. A car that averaged $50\ \text{km/h}$ over an hour may have sat at a red light for some of it and sped on the highway for the rest.

Shrinking the interval

To find the velocity at one particular instant, we shrink the interval. Keep the first point fixed and slide the second point toward it. As Δt gets smaller, the secant line pivots, and it settles on a single line that just grazes the curve at the point of interest: the **tangent line** (Figure D.1). Another way to see it: zoom in on any smooth curve far enough and it looks straight. The slope of that nearly straight piece is the rate of change at that instant.

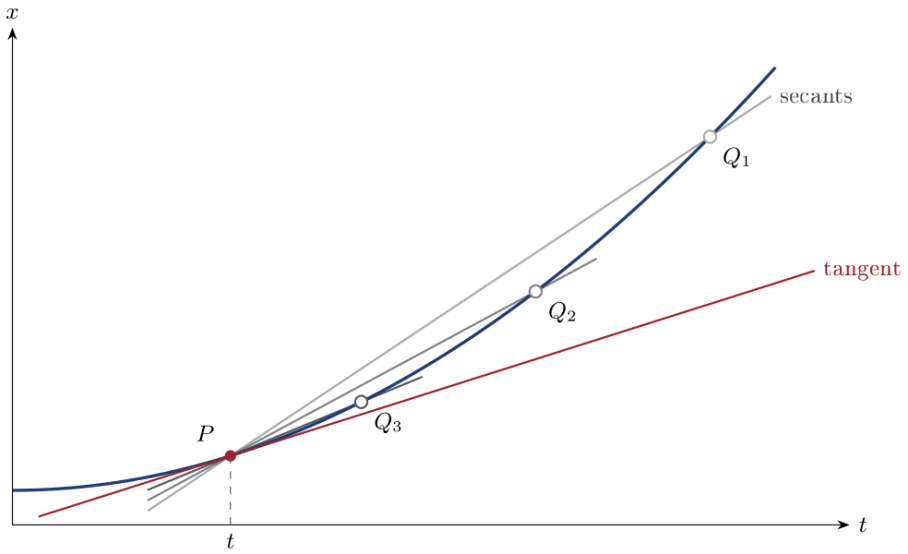


Figure D.1. From average rate to instantaneous rate. Each gray secant line joins P to a later point (Q_1, Q_2, Q_3) on the position curve; its slope is the average velocity over that interval. As the second point slides toward P, the secants pivot toward the red tangent line, whose slope is the instantaneous velocity at P.

•Figure description

A wide graph of position x versus time t showing a curve that bends gently upward. A point P is marked on the curve at an early time t. Three gray secant lines run from P to later points on the curve, labeled Q1, Q2 and Q3, with Q3 closest to P; the closer the second point is to P, the less steep its secant. A red line, the tangent, touches the curve only at P and lies below the curve everywhere else. As the second point moves from Q1 toward P, the secants pivot down toward the tangent, showing that the slope of the tangent is the limit of the average slopes.

We write the result of this limiting process as

Key Equation: The derivative

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}. \tag{D.1}$$

This is the **derivative** of x with respect to t. Read the symbol d as "a tiny change in." Then dx/dt says just what it looks like: a tiny change in x divided by the tiny change in t that produced it. It is a ratio of two very small but real quantities, which is why it has units. The units of any derivative are the units of the top quantity divided by the units of the bottom one, so dx/dt is in $\mathrm{m/s}$. This is the instantaneous velocity of equation (2.4).

Watching the limit happen

There is nothing abstract about this limit; you can watch it happen with a calculator. Let an object's position be $x(t) = (5.0\ \mathrm{m/s^2})\ t^2$, and ask for its velocity at $t = 2.0\ \mathrm{s}$, where $x = 20\ \mathrm{m}$. Compute the average velocity from $t = 2.0\ \mathrm{s}$ to $t = 2.0\ \mathrm{s} + \Delta t$ for smaller and smaller Δt:

Average velocity over shrinking time intervals

Δt (s)	x at $2 + \Delta t$ (m)	Δx (m)	$\Delta x/\Delta t$ (m/s)
1	45	25	25
0.1	22.05	2.05	20.5
0.01	20.2005	0.2005	20.05
0.001	20.020005	0.020005	20.005

The averages close in on $20\ \mathrm{m/s}$. Algebra shows why. For any t,

$$\frac{\Delta x}{\Delta t} = \frac{5(t + \Delta t)^2 - 5t^2}{\Delta t} = \frac{10\,t\,\Delta t + 5(\Delta t)^2}{\Delta t} = 10\,t + 5\,\Delta t. \tag{D.1.2}$$

As $\Delta t \to 0$ the last term disappears, leaving $dx/dt = 10\ t$, which is $20\ \mathrm{m/s}$ at $t = 2.0\ \mathrm{s}$. The leftover $5\ \Delta t$ is exactly the error in the table: half a meter per second when $\Delta t = 0.1$, a twentieth of that when $\Delta t = 0.01$.

What the slope tells you

Because a derivative is a slope, its sign and size can be read straight off a graph:

- Where the graph rises, the slope is positive and the quantity is increasing.
- Where the graph falls, the slope is negative and the quantity is decreasing.
- Where the graph is momentarily flat, at a peak or a valley, the slope is zero. This is how physics finds a maximum or a minimum: the top of a thrown ball's flight is the instant its height stops increasing, $dy/dt = 0$.
- Where the graph is steep, the quantity is changing quickly; where it is shallow, slowly.

Chapter 2 uses exactly this reasoning to read velocity off a position graph and acceleration off a velocity graph (Figure 2.3).

A derivative is any rate

Nothing in the definition requires x to be a position or t to be a time. Any quantity that depends on another has a rate of change with respect to it. The derivatives you will meet in this book include:

Rates of change used in this book

Rate	Meaning	Where
$v = dx/dt$	how fast position changes	Eq. (2.4)

Rate	Meaning	Where
$a = dv/dt$	how fast velocity changes	Eq. (2.6)
$P = dW/dt$	how fast work is done	Eq. (7.9)
$\mathbf{F}_{\text{net}} = d\mathbf{p}/dt$	how fast momentum changes	Eq. (9.3)
$I = dQ/dt$	how fast charge flows past a point	Eq. (20.1)
$E_x = -\partial V/\partial x$	how fast potential changes with *position*	Eq. (18.16)
$\mathcal{E} = -d\Phi_B/dt$	how fast magnetic flux changes	Eq. (24.3)

The electric field entry is a rate with respect to distance rather than time; the curly ∂ only signals that V depends on more than one coordinate and we are letting just one of them change. The idea is the same in every row: a small change in one thing divided by the small change in another.

A few derivatives worth knowing

You can always find a derivative by computing $\Delta x/\Delta t$ and shrinking Δt, as we did above. In practice, the results for a handful of functions cover nearly everything in this book. Here C and n are constants and u is the variable.

Derivatives of common functions

Function	Derivative with respect to u
C (a constant)	0
u^n	$n\ u^{n-1}$
$\sin(ku)$	$k\cos(ku)$
$\cos(ku)$	$-k\sin(ku)$
e^{ku}	$k\ e^{ku}$
$\ln u$	$1/u$
$C\ f(u)$	$C\ \mathrm{d}f/\mathrm{d}u$
$f(u) + g(u)$	$\mathrm{d}f/\mathrm{d}u + \mathrm{d}g/\mathrm{d}u$

The trigonometric rules hold only when angles are in radians, one more reason physics measures angles that way (Chapter 10). When a function sits inside another, as in $\sin(ku)$ or e^{ku}, the derivative picks up the rate of the inside function as a factor: here, the k. That is the **chain rule**: the rates multiply.

Example D.1: The top of the flight

A ball is thrown straight up with initial speed $v_0 = 14.7\ \text{m/s}$. Its height is $y(t) = v_0\ t - \frac{1}{2}g\ t^2$, with $g = 9.80\ \text{m/s}^2$. Find its velocity and acceleration as functions of time, and the greatest height it reaches.

Solution

Differentiate term by term. The derivative of t is 1 and the derivative of t^2 is $2t$, so

$$v = \frac{\mathrm{d}y}{\mathrm{d}t} = v_0 - g\,t, \qquad a = \frac{\mathrm{d}v}{\mathrm{d}t} = -g. \tag{D.1.3}$$

The acceleration is constant: free fall. The ball is highest where the height graph is flat, where $v = 0$:

$$t_{\text{top}} = \frac{v_0}{g} = \frac{14.7\,\text{m/s}}{9.80\,\text{m/s}^2} = 1.50\,\text{s}, \qquad y_{\max} = (14.7)(1.50) - \tfrac{1}{2}(9.80)(1.50)^2 = 11.0\,\text{m}. \tag{D.1.4}$$

Check: before t_{top}, $v > 0$ and the ball rises; after it, $v < 0$ and the ball falls. The slope changes sign at the peak, exactly as it should.

Checkpoint D.1

A position–time graph is a straight line sloping downward. What can you say about the object's velocity? What about its acceleration?

Answers to Checkpoints

Checkpoint D.1. The slope is constant and negative, so the velocity is constant and negative: the object moves in the $-x$ direction at a steady speed. Since the velocity does not change, the acceleration is zero.

This page titled D.1: The Derivative: Slope as a Rate of Change is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.2: The Integral: Adding Up Small Pieces

Now turn the dashboard question around. Suppose the odometer is broken, but you have a record of the speedometer: the velocity v at every instant. How far did the car go?

If the velocity were constant, the answer would be easy: distance equals velocity times time, $\Delta x = v\ \Delta t$. On a graph of v versus t, that product is the area of a rectangle of height v and width Δt. The displacement is the area under the velocity graph.

If the velocity changes, no single rectangle will do. But over a short enough interval, the velocity hardly changes at all. So we *chop the trip into many short intervals*, each of length Δt. Over interval number i the velocity is nearly some constant v_i, so the car moves about $v_i\ \Delta t$, the area of one thin strip. Adding up all the strips gives the total:

$$\Delta x \approx \sum_i v_i\, \Delta t. \tag{D.2.1}$$

A sum like this is called a **Riemann sum**. It is only approximate, because v does change a little inside each strip. But the error shrinks as the strips get thinner, and in the limit of infinitely many infinitely thin strips the sum becomes exact (Figure D.2).

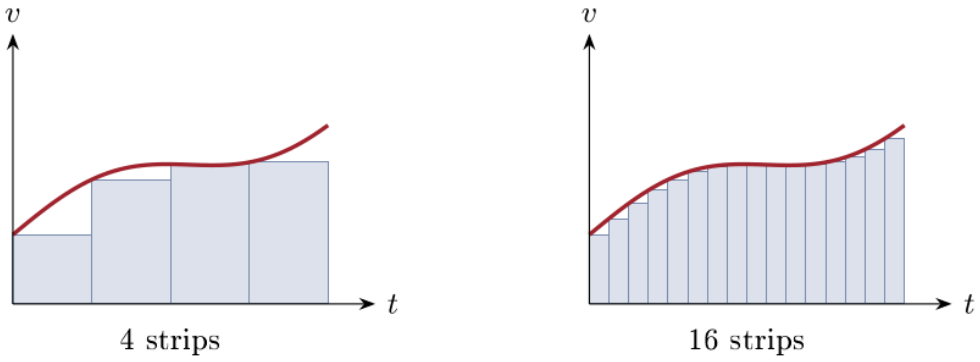


Figure D.2. Integration as a sum of thin strips. Each strip has width Δt and height equal to the velocity at its left edge, so its area is approximately the displacement during that short interval. With four strips (left) the approximation misses noticeably; with sixteen (right) the error is already small. In the limit of infinitely many, infinitely thin strips, the sum equals the area under the curve exactly.

•Figure description

Two side-by-side graphs of the same smooth, rising curve of velocity v versus time t. In the left graph the region under the curve is approximated by four wide rectangles whose tops are set by the curve's height at each rectangle's left edge; large uncovered gaps are visible between the rectangle tops and the curve. In the right graph the same region is approximated by sixteen narrow rectangles, and the gaps between them and the curve are much smaller. The figure shows that thinner strips make the sum of rectangle areas a better approximation to the area under the curve.

Watching the sum converge

Here, too, you can watch the limit happen. Let $v = (10\ \text{m/s}^2)\ t$ from $t = 0$ to $t = 3.0\ \text{s}$. The graph is a straight line, so the exact area is a triangle: $\frac{1}{2}(3.0\ \text{s})(30\ \text{m/s}) = 45\ \text{m}$. Now pretend we did not know that and add up N strips, each with its height taken at its left edge:

Riemann sums for the displacement, with more and more strips

Number of strips N	Strip width Δt (s)	$\sum v_i\ \Delta t$ (m)
3	1.0	30.0
6	0.5	37.5
30	0.1	43.5
300	0.01	44.85

(With $N = 3$, for example, the heights are $v = 0$, 10, and $20\ \mathrm{m/s}$, and $(0 + 10 + 20)(1.0) = 30$.) The sums approach $45\ \mathrm{m}$. They all fall short because the velocity is increasing, so each strip's left-edge height is the lowest value in its interval. Every strip misses a small triangle at its top, and those triangles shrink faster than the number of strips grows.

The integral sign

The limit of the sum is written with a new symbol:

Key Equation: The integral as a sum

$$\int_{t_i}^{t_f} v\,dt \;=\; \lim_{\Delta t \to 0} \sum_i v_i\,\Delta t. \tag{D.2}$$

This is the **integral** of v with respect to t from t_i to t_f. The sign $\int$ is a stretched letter S, for *sum*. Every piece of the notation has a plain meaning:

- dt is the width of one small piece. It is not just notation: like Δt, it is an actual stretch of time, measured in seconds, only short enough that v has no chance to change within it. A quantity this small is called **infinitesimal**.
- v, the **integrand**, is "how much per unit of t" for the piece at time t. The product $v\ dt$ is the small contribution from one piece.
- t, the **variable of integration**, *labels* the pieces. Each piece has its own value of t, and that is how we tell them apart.
- The **limits of integration** t_i and t_f mark where the adding starts and where it stops.

The units of an integral are the units of the integrand times the units of the small piece: $(\ \mathrm{m/s})(\ \mathrm{s}) =\ \mathrm{m}$ here. Whenever an integral appears, you can read it aloud as "add up all the little $v\ dt$'s from t_i to t_f."

Area is just one picture

Graphically, an integral is the **area under a curve**, provided we count area below the horizontal axis as negative. That is a useful picture: Chapter 2 reads displacement as the signed area under a velocity graph (Figure 2.4), and Chapter 7 reads work as the area under a force–position graph.

But the deeper idea is the sum, not the area. The area under a graph is seldom the thing a physicist is actually after. We integrate to find the *total* of something by adding up contributions from a great many tiny parts: the mass of a rod from the masses of its slices, the work done by a varying force from the work done over each small step, the electric field of a charged object from the fields of its tiny bits of charge. Whenever you see $\int$, think: *we are adding up small pieces of something to get the whole*. Hold on to that reading and most integrals in physics can be understood before a single one is evaluated.

Checkpoint D.2

(a) If F is a force in newtons and t a time in seconds, what are the units of $\int F\ dt$? What physical quantity has those units?
(b) If Q is a charge in coulombs, what are the units of dQ/dt?

Answers to Checkpoints

Checkpoint D.2. (a) $\mathrm{N \cdot s = kg \cdot m/s}$, the units of momentum; $\int F\ dt$ is the impulse. (b) $\mathrm{C/s}$, which is the ampere; dQ/dt is an electric current.

This page titled D.2: The Integral: Adding Up Small Pieces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.3: Why the Two Ideas Undo Each Other

We could always evaluate an integral by adding up thousands of strips on a computer, and sometimes physicists do exactly that. But there is a far quicker way, and it comes from the dashboard picture.

Let $A(t)$ be the area under the velocity graph from the start of the trip up to time t: the odometer reading. How fast does A grow? Move forward by a tiny time dt and you add one more thin strip, of height $v(t)$ and width dt (Figure D.3). So $dA = v(t)\ dt$, or

$$\frac{\mathrm{d}A}{\mathrm{d}t} = v(t). \tag{D.3.1}$$

The rate at which the area grows is the height of the curve. The odometer's reading changes at the rate shown on the speedometer.

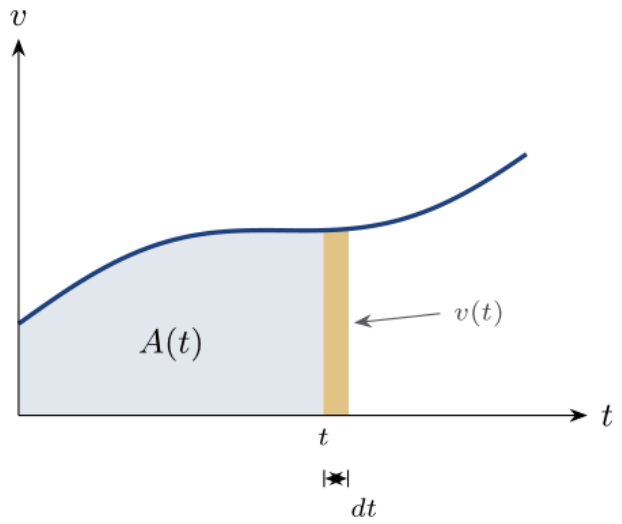


Figure D.3. Why integration undoes differentiation. The shaded region is the area $A(t)$ accumulated up to time t. Advancing by dt adds one thin strip of area $v(t)\ dt$, so A grows at the rate $dA/dt = v(t)$.

•Figure description

A graph of a smooth rising curve v versus t. The region under the curve from time zero up to a time labeled t is shaded light blue and labeled A of t, the area accumulated so far. Immediately to its right, a single narrow strip from t to t plus d t is shaded darker; its height reaches the curve and is labeled v of t, and its width is labeled d t. The figure shows that advancing time by d t adds an area of v times d t, so the accumulated area grows at a rate equal to the curve's height.

This gives the recipe for evaluating an integral. To integrate v, find any function whose derivative is v. Such a function is called an **antiderivative** of v. Then the area between t_i and t_f is the change in that function between those two times:

Key Equation: Fundamental theorem of calculus

$$\int_a^b f(u)\,du = F(b) - F(a), \qquad \text{where } \frac{\mathrm{d}F}{\mathrm{d}u} = f(u). \tag{D.3}$$

This is the **fundamental theorem of calculus**. It says that the two questions we started with are really one question asked in opposite directions. Differentiation turns a total into a rate; integration turns a rate back into a total. The vertical bar notation $F(u)\big|_a^b$ is shorthand for $F(b) - F(a)$.

Return to the strips of the last section. An antiderivative of $v = 10\ t$ is $5\ t^2$, since the derivative of $5\ t^2$ is $10\ t$. So

$$\int_0^{3.0} 10\,t\,dt = 5\,t^2\Big|_0^{3.0} = 5(3.0)^2 - 5(0)^2 = 45\,\mathrm{m}, \tag{D.3.2}$$

the value the strips were creeping toward, found in one line.

Reading the derivative table of Section D.1 backward gives the integrals you will need most often:

Common antiderivatives

Integrand	Antiderivative
C (a constant)	$C\ u$
$u^n \quad (n \neq -1)$	$\dfrac{u^{n+1}}{n+1}$
$1/u$	$\ln u$
$\cos(ku)$	$\frac{1}{k}\sin(ku)$
$\sin(ku)$	$-\frac{1}{k}\cos(ku)$
e^{ku}	$\frac{1}{k}\ e^{ku}$

Constants come outside an integral, and the integral of a sum is the sum of the integrals. Harder integrals can be looked up in a table or done by software; in physics the challenge is almost never the integration itself. It is *setting up* the integral, which is the subject of the next section.

Note: One d, one meaning

The d in a derivative and the d in an integral mean the same thing: "a tiny amount of." That is why physicists happily treat dx and dt as ordinary quantities. If $v = dx/dt$, then $dx = v\ dt$: the small distance traveled is the velocity times the small time. If the charge per unit length is $\lambda = dq/dx$, then $dq = \lambda\ dx$. Writing the small piece this way is the first step of almost every integral in this book.

Checkpoint D.3

For part of a trip, a car's speedometer reads zero. During that time, what does the graph of its odometer reading look like? Later the car backs up, so its velocity is negative. What does the accumulated displacement $A(t)$ do then?

Answers to Checkpoints

Checkpoint D.3. While the speedometer reads zero, the odometer graph is flat: zero slope, no change. While the velocity is negative, each new strip has negative area, so the accumulated displacement $A(t)$ decreases. (An odometer, which records distance rather than displacement, would keep increasing.)

This page titled D.3: Why the Two Ideas Undo Each Other is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.4: Integration in Physics: Slice, Label, Add

Most integrals in physics are built the same way, and once you recognize the pattern they become routine:

1. **Set up a coordinate system.** You need a way to say where each piece is.
2. **Slice the object into small pieces** and label each piece by a single coordinate. Choose the slices so that everything you need (density, distance, direction) is the same throughout any one slice.
3. **Write the small amount** of the quantity you want, contributed by one slice, in terms of that coordinate and its d: for example $dm = \lambda\ dx$.
4. **Add up the slices** with an integral. The limits are the coordinates of the first and last slices.
5. **Check the answer** with units, a limiting case, or a rough average.

The examples below use this recipe on mass and area, where the answers are easy to check. Chapters 15 and 16 use the identical recipe to find forces and fields from continuous charge distributions (Section 15.4).

Example D.2: Mass of a nonuniform rod

A thin rod lies along the x-axis from $x = 0$ to $x = 2.0\ \mathrm{m}$. It is made thicker toward one end, so its linear mass density increases along its length as

$$\lambda(x) = 0.20\,\mathrm{kg/m} + (0.30\,\mathrm{kg/m^2})\,x. \tag{D.4.1}$$

Find the rod's total mass.

Solution

If the density were uniform we would just multiply, $m = \lambda L$. It is not, so we slice. Cut the rod into thin slices of length dx; the coordinate x labels each slice. A slice is so thin that λ is constant across it, so its mass is "mass per length times length":

$$dm = \lambda\,dx. \tag{D.4.2}$$

The first slice is at $x = 0$ and the last at $x = 2.0\ \mathrm{m}$, so the total mass is

$$m = \int dm = \int_0^{2.0} \left(0.20 + 0.30\,x\right) dx = \Big[0.20\,x + 0.15\,x^2\Big]_0^{2.0} = 0.40 + 0.60 = 1.0\,\mathrm{kg}. \tag{D.4.3}$$

Check: the density runs from $0.20\ \mathrm{kg/m}$ at one end to $0.80\ \mathrm{kg/m}$ at the other. Since it varies linearly, its average is the midpoint value, $0.50\ \mathrm{kg/m}$, and $(0.50\ \mathrm{kg/m})(2.0\ \mathrm{m}) = 1.0\ \mathrm{kg}$. The integral and the average agree. For a density that is not linear the simple average will no longer be exact, but it remains an excellent way to catch a large mistake.

Choosing the slices: the area of a circle

The same recipe finds areas and volumes. To find the area of a circle of radius R, slice it into thin concentric rings, like the grooves on a vinyl record (Figure D.4). The ring at radius r has width dr, and r labels each ring. Imagine cutting one ring and straightening it out: it becomes a thin strip of length $2\pi r$ (the ring's circumference) and width dr. So its area is length times width,

$$dA = 2\pi r\,dr. \tag{D.4.4}$$

The rings run from $r = 0$ at the center to $r = R$ at the edge:

$$A = \int dA = \int_0^R 2\pi r\,dr = 2\pi\,\frac{r^2}{2}\Big|_0^R = \pi R^2, \tag{D.4.5}$$

the familiar result. Building a ball out of thin spherical *shells*, like the layers of an onion, works the same way in three dimensions: a shell of radius r and thickness dr has volume $dV = 4\pi r^2\ dr$ (its surface area times its thickness), and $\int_0^R 4\pi r^2\ dr = \frac{4}{3}\pi R^3$, the volume of a sphere.

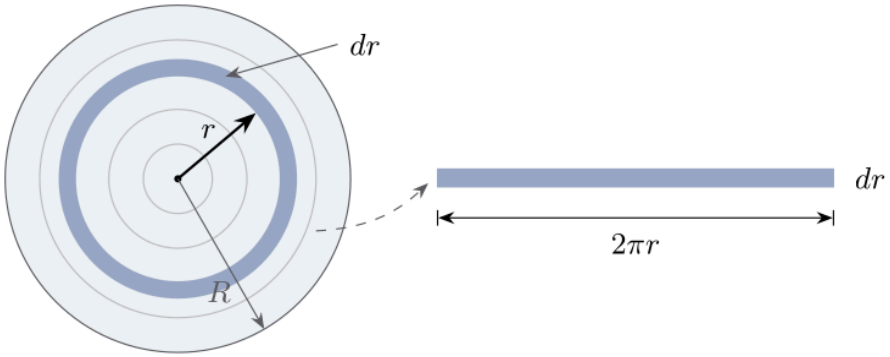


Figure D.4. Slicing a disk into rings. A thin ring at radius r with width dr, cut and straightened, is a strip of length $2\pi r$ and width dr, so its area is $dA = 2\pi r\ dr$. Adding the rings from $r = 0$ to $r = R$ gives πR^2.

•Figure description

On the left, a disk of radius R with one thin ring at radius r highlighted. An arrow from the center marks the radius r, and the ring's small width is labeled d r. On the right, the same ring is shown cut and straightened into a long thin rectangle whose length is labeled two pi r and whose width is labeled d r. The figure shows that the ring's area is approximately its circumference times its width.

Why rings rather than, say, vertical strips? Both would work for the area alone, but rings make the integral simple because every point of a ring is the same distance from the center. That matters as soon as something depends on r, as in the next example. The general rule: *choose slices over which everything in the integrand is constant*. For a cylinder whose density changes along its length, slice it crosswise into thin disks of thickness dz. For one whose density changes with distance from its axis, slice it into nested cylindrical shells of thickness dr.

Example D.3: A disk thinning toward the rim

A thin disk of radius R is thickest at its center and tapers to nothing at its edge, so its surface mass density falls off linearly with distance from the center: $\sigma(r) = \sigma_0\ (1 - r/R)$. It equals σ_0 at the center and zero at the rim. Find the mass of the disk.

Solution

The density depends on r, so slice into rings: every point of a ring has the same density. The ring at radius r has area $dA = 2\pi r\ dr$, so its mass is "mass per area times area,"

$$dm = \sigma\, dA = \sigma_0\Big(1 - \frac{r}{R}\Big)\, 2\pi r\, dr = 2\pi\sigma_0\Big(r - \frac{r^2}{R}\Big)dr. \tag{D.4.6}$$

Add the rings from the center to the rim:

$$m = \int_0^R 2\pi\sigma_0\Big(r - \frac{r^2}{R}\Big)dr = 2\pi\sigma_0\Big(\frac{R^2}{2} - \frac{R^2}{3}\Big) = \tfrac{1}{3}\,\sigma_0\,\pi R^2. \tag{D.4.7}$$

Check: the mass is the same as for a uniform disk with density $\frac{1}{3}\sigma_0$. That is sensible. The density runs from σ_0 down to 0, so the effective average must lie between them, and it should lie below the midpoint $\frac{1}{2}\sigma_0$ because most of a disk's area is near the rim, where the density is low. Had we found an average density larger than σ_0, or negative, we would know we had made a mistake.

Checkpoint D.4

A long solid cylinder carries charge whose volume density depends only on the distance from its axis. To find its total charge, should you slice it into disks of thickness dz or into cylindrical shells of thickness dr? Why?

When the pieces are vectors

Sometimes each small piece contributes a *vector*: a small force or a small electric field, each pointing in its own direction. You

cannot add vectors by adding their magnitudes, but you can add their components. So a vector integral is really two or three ordinary integrals, one for each component, often with some of them canceling by symmetry. Chapters 15 and 16 work through this step by step, beginning with equations (15.6) and (16.8). The recipe is the one above; only the bookkeeping is longer.

Integrals you will meet in this book

Every integral below means the same thing: add up the small pieces to get the whole.

Integrals used in this book

Integral	Adds up	Where
$\Delta x = \int v\ dt$	small displacements $v\ dt$	Ch. 2
$W = \int F\ dx$	work over small steps $F\ dx$	Eq. (7.6)
$\mathbf{J} = \int \mathbf{F}\ dt$	small momentum changes $\mathbf{F}\ dt$	Eq. (9.4)
$I = \int r^2\ dm$	each bit of mass times r^2	Eq. (10.10)
$\mathbf{E} = \int d\mathbf{E}$	fields of small charges dq	Eq. (16.8)
$\Delta V = -\int \mathbf{E} \cdot d\mathbf{s}$	potential changes over small steps	Eq. (18.1)
$\Phi_B = \int \mathbf{B} \cdot d\mathbf{A}$	flux through small patches $d\mathbf{A}$	Eq. (24.2)

Answers to Checkpoints

Checkpoint D.4. Cylindrical shells. Every point of a thin shell is the same distance from the axis, so the density is constant across the shell and its charge is simply $dq = \rho\ (2\pi r L)\ dr$. Across a disk, by contrast, the density would vary from center to edge.

This page titled D.4: Integration in Physics: Slice, Label, Add is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.5: Checking a Calculus Result

Calculus gives exact answers, but it is easy to drop a factor or a sign on the way. A few quick checks catch most errors, and none of them needs any calculus:

- **Units.** A derivative has the units of the top over the units of the bottom; an integral has the units of the integrand times the units of $d(\text{variable})$. If the units come out wrong, so does the answer.
- **Sign.** Is the quantity increasing or decreasing? Is the area above or below the axis? The sign of the answer should match.
- **Size, by averaging.** A well-chosen average value times the size of the interval should land close to the integral, as in Examples D.2 and D.3. The average must lie between the smallest and largest values of the integrand.
- **A special case you already know.** Set the varying quantity to a constant, or let a distance become very large, and see whether a familiar result appears. A uniform rod should give $m = \lambda L$; a charge distribution seen from far away should look like a point charge.

This page titled D.5: Checking a Calculus Result is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.S: Motivation of Calculus for Physics Students (Summary)

Summary

- Calculus answers two questions: how fast a quantity is changing (the derivative) and how much has built up in total (the integral).
- The **derivative** $dx/dt = \lim_{\Delta t \to 0} \Delta x/\Delta t$ is a rate of change. On a graph it is the slope of the tangent line: positive where the graph rises, negative where it falls, zero at a peak or valley. Its units are the units of x divided by the units of t.
- The **integral** $\int_a^b f\ du = \lim \sum f_i\ \Delta u$ is a sum of many small pieces. On a graph it is the signed area under the curve. Its units are the units of f times the units of u.
- In $\int f\ du$, du is the size of one small piece, u labels the pieces, $f\ du$ is one piece's contribution, and the limits mark the first and last pieces.
- Differentiation and integration undo each other (the fundamental theorem of calculus): to integrate f, find a function F with $dF/du = f$ and evaluate $F(b) - F(a)$.
- To set up a physics integral: choose coordinates, slice so that everything is constant within a slice, write the small piece ($dm = \lambda\ dx$, $dA = 2\pi r\ dr$, ...), add with the right limits, and check.

Key Equations

The derivative (a rate of change; slope of the tangent)

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \tag{D.S.1}$$

The integral (a sum of small pieces; area under the curve)

$$\int_{t_i}^{t_f} v\, dt = \lim_{\Delta t \to 0} \sum_i v_i\, \Delta t \tag{D.S.2}$$

Fundamental theorem of calculus

$$\int_a^b f(u)\, du = F(b) - F(a), \qquad \frac{\mathrm{d}F}{\mathrm{d}u} = f(u) \tag{D.S.3}$$

Small pieces used most often

$$dm = \lambda\, dx, \qquad dA_{\text{ring}} = 2\pi r\, dr, \qquad dV_{\text{shell}} = 4\pi r^2\, dr \tag{D.S.4}$$

This page titled D.S: Motivation of Calculus for Physics Students (Summary) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

D.E: Motivation of Calculus for Physics Students (Practice Problems)

Answers follow the problems.

Derivatives as Rates

D.1 ★ An object's position is $x(t) = (4.0\ \text{m/s}^3)\ t^3 - (2.0\ \text{m/s})\ t$. Find its velocity and acceleration at $t = 2.0\ \text{s}$.

D.2 ★ A cup of coffee cools according to $T(t) = 20\,^\circ\text{C} + (70\,^\circ\text{C})\ e^{-t/(10\ \text{min})}$. (a) What is dT/dt, and what are its units? (b) How fast is the coffee cooling at $t = 0$ and at $t = 10\ \text{min}$? (c) Explain from the shape of the graph why the second rate is smaller.

D.3 ★★ Following the table in Section D.1, compute the average velocity of $x(t) = (5.0\ \text{m/s}^2)\ t^2$ between $t = 1.0\ \text{s}$ and $t = 1.0\ \text{s} + \Delta t$ for $\Delta t = 0.1$, 0.01, and $0.001\ \text{s}$. What value do the averages approach, and does it agree with $dx/dt = 10\ t$?

Integrals as Sums and Areas

D.4 ★ An object's velocity is $v(t) = (2.0\ \text{m/s}^3)\ t^2$. (a) Estimate its displacement from $t = 0$ to $t = 3.0\ \text{s}$ using three rectangles of width $1.0\ \text{s}$, each with its height taken at its left edge. (b) Find the exact displacement by integrating. (c) Explain why your estimate is too small.

D.5 ★ Find the area under one arch of the curve $y = \sin x$, from $x = 0$ to $x = \pi$. Before integrating, estimate the answer: the arch fits inside a rectangle of width π and height 1, so the area must be less than π.

D.6 ★ A force that varies with position, $F(x) = (3.0\ \text{N/m}^2)\ x^2$, pushes an object from $x = 0$ to $x = 2.0\ \text{m}$. How much work does it do?

D.7 ★★ The current in a discharging circuit is $I(t) = I_0\ e^{-t/\tau}$. How much charge flows past a point from $t = 0$ until the current has died away completely ($t \to \infty$)? Give the answer in terms of I_0 and τ, and check its units.

Slicing Physical Objects

D.8 ★ A rod lies along the x-axis from $x = 0$ to $x = L$ and carries charge with linear density $\lambda(x) = \lambda_0\ x^2/L^2$. Find the total charge on the rod. Then check your answer: should it be more or less than $\frac{1}{2}\lambda_0 L$?

D.9 ★★ A bowl is a hemisphere of radius R. Find its volume by slicing it into thin horizontal disks of thickness dz. (Hint: a disk a height z above the flat face has radius $r = \sqrt{R^2 - z^2}$.) Check your answer against the volume of a whole sphere.

D.10 ★★ A solid sphere of radius R has a charge density that grows linearly from zero at the center, $\rho(r) = \rho_0\ r/R$. Slice it into thin spherical shells and find its total charge. Express your answer as a uniform-sphere result with an effective average density, and compare that average to ρ_0.

D.11 ★★★ A uniform thin rod has mass M and length L. Its moment of inertia about an axis is $I = \int r^2\ dm$, where r is each piece's distance from the axis (equation (10.10)). (a) Slice the rod into pieces of length dx and find I about an axis through one end. (b) Repeat for an axis through the center. (c) Explain physically why the two answers differ.

Answers to Practice Problems

D.1 $v = 12t^2 - 2 = 46\ \text{m/s}$; $a = 24t = 48\ \text{m/s}^2$.

D.2 (a) $dT/dt = -(7.0\ ^\circ\text{C/min})\ e^{-t/(10\ \text{min})}$, in $^\circ\text{C/min}$. (b) $-7.0\ ^\circ\text{C/min}$ at $t = 0$; $-2.6\ ^\circ\text{C/min}$ at $t = 10\ \text{min}$. (c) The graph flattens as the coffee approaches room temperature, so its slope shrinks.

D.3 $10.5\ \text{m/s}$, $10.05\ \text{m/s}$, $10.005\ \text{m/s}$; they approach $10\ \text{m/s} = 10(1.0)$, in agreement.

D.4 (a) $(0 + 2.0 + 8.0)(1.0) = 10\ \text{m}$. (b) $\frac{2.0}{3}(3.0)^3 = 18\ \text{m}$. (c) The velocity increases throughout, so each left-edge height is the smallest value in its interval.

D.5 $\left[-\cos x\right]_0^{\pi} = 2$, about 64% of the enclosing rectangle.

D.6 $W = \frac{3.0}{3}(2.0)^3 = 8.0\ \text{J}$.

D.7 $Q = I_0\ \tau$; amperes times seconds is coulombs.

D.8 $Q = \frac{1}{3}\lambda_0 L$; less than $\frac{1}{2}\lambda_0 L$, because the density is small over most of the rod and large only near $x = L$.

D.9 $V = \int_0^R \pi(R^2 - z^2)\ dz = \frac{2}{3}\pi R^3$, half of $\frac{4}{3}\pi R^3$.

D.10 $Q = \pi\rho_0 R^3 = \frac{3}{4}\rho_0 \cdot \frac{4}{3}\pi R^3$; the effective average density is $\frac{3}{4}\rho_0$, between 0 and ρ_0 and weighted toward the outside, where most of the volume is.

D.11 Each piece has mass $dm = (M/L)\ dx$. (a) $I = \int_0^L x^2\ (M/L)\ dx = \frac{1}{3}ML^2$. (b) $I = \int_{-L/2}^{L/2} x^2\ (M/L)\ dx = \frac{1}{12}ML^2$. (c) About the center, no mass is farther than $L/2$ from the axis, and because r is squared, the distant mass dominates.

This page titled D.E: Motivation of Calculus for Physics Students (Practice Problems) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Michael Ragsdale.

INDEX

A

absolute zero
36.1: Temperature

absorption (of light)
41.7: Lasers

absorption spectrum
41.3: The Hydrogen Atom

acceleration
2.4: Motion Graphs

acceleration, angular
10.1: Rotational Kinematics

acceleration, average
2.S: Motion in One Dimension (Summary)

acceleration, centripetal
4.3: Uniform Circular Motion

acceleration, constant
2.S: Motion in One Dimension (Summary)

acceleration, instantaneous
2.S: Motion in One Dimension (Summary)

acceptor (semiconductor)
42.4: Semiconductors

accommodation
32.5: Optical Instruments

accuracy
1.S: Measurement (Summary)

action at a distance
16: Electric Fields

action–reaction pair
5.1: Newton's Laws of Motion

activity (radioactive)
43.3: Radioactivity

adiabatic index
37.3: Thermodynamic Processes

age of the universe
44.4: Cosmology

Airy disk
34.2: Diffraction by a Circular Aperture

alpha decay
43.4: Nuclear Reactions

alternating current
26.2: Resistors in an AC Circuit

ammeter
21.1: Circuit Elements and Diagrams

ampere (unit)
20.1: Electric Current

ampere (unit), definition of
23.2: Force Between Two Parallel Currents

Ampère–Maxwell law
27.1: Displacement Current

Ampère's law
23.4: Ampère's Law
27.2: Maxwell's Equations

Ampèrian loop
23.4: Ampère's Law

amplitude
28.3: Mass–Spring System

angular acceleration
10.1: Rotational Kinematics

angular displacement
10.1: Rotational Kinematics

angular frequency
28.1: Simple Harmonic Motion
29.2: The Wave Equation

angular magnification
32.5: Optical Instruments

angular momentum
11: Rotation II

angular momentum, orbital
41.5: The Quantum Model of the Hydrogen Atom

angular momentum, spin
41.6: The Pauli Exclusion Principle

angular velocity
10.1: Rotational Kinematics
10.4: Torque and the Right-Hand Rule

annihilation
44.1: Elementary Particles

antenna
27.6: The Electromagnetic Spectrum

anti-reflection coating
33.3: Interference from Thin Films

antiderivative
D.3: Why the Two Ideas Undo Each Other

antimatter
44.1: Elementary Particles

antinode
29.6: Standing Waves and Resonance

aphelion
13.4: Kepler's Laws of Planetary Motion

apparent weight
5.5: Applying Newton's Laws

Archimedes' principle
14.4: Buoyancy and Archimedes' Principle

area under a curve
2.S: Motion in One Dimension (Summary)
D.2: The Integral: Adding Up Small Pieces

area vector
17.2: Electric Flux

arrow of time
38.2: The Second Law of Thermodynamics

atomic mass unit
43.1: Properties of Nuclei

atomic model, billiard-ball
41.1: Models of the Atom

atomic model, Bohr
41.1: Models of the Atom

atomic model, nuclear
41.1: Models of the Atom

atomic model, plum-pudding
41.1: Models of the Atom

atomic model, quantum
41.1: Models of the Atom

atomic number
43.1: Properties of Nuclei

atomic planes
34.5: Diffraction of X-Rays by Crystals

aurora
22.2: Motion of a Charged Particle in Magnetic Fields

average acceleration
2.S: Motion in One Dimension (Summary)

average velocity
2.S: Motion in One Dimension (Summary)

Avogadro's number
36.3: Ideal Gas Law

B

back emf
24.4: Generators and Motors

ballistic pendulum
9.E: Linear Momentum and Center of Mass (Problems)

band gap
42.3: Electrical Conduction in Metals and Insulators

band theory
42.3: Electrical Conduction in Metals and Insulators

baryon
44.1: Elementary Particles

baryon number
44.1: Elementary Particles

base of support
12.2: Center of Gravity and Stability

base units
1.2: Measuring Standards

basis vectors, standard
3.3: Unit Vectors

battery
21.1: Circuit Elements and Diagrams

battery, ideal
21.1: Circuit Elements and Diagrams

battery, real
21.4: Ideal vs. Real Batteries

beam splitter
33.4: Interferometry

beat frequency
30.5: Beats

beats
30.5: Beats

becquerel
43.3: Radioactivity

Bernoulli's equation
14.5: Fluid Dynamics and Bernoulli's Principle

beta decay
43.4: Nuclear Reactions

Big Bang
44.4: Cosmology

bimetallic strip
36.2: Thermal Expansion

binding energy
43.2: Nuclear Binding Energy

binding energy, per nucleon
43.2: Nuclear Binding Energy

Biot–Savart law
23.1: Magnetic Field Due to a Current

black hole
35.6: General Relativity
blackbody
39.1: Blackbody Radiation
44.4: Cosmology
blackbody radiation
39.1: Blackbody Radiation
blueshift
35.4: Doppler Effect for Light
Bohr magneton
23.7: Magnetic Properties of Matter
Bohr model
41.1: Models of the Atom
Bohr radius
41.3: The Hydrogen Atom
Bohr, Niels
41.1: Models of the Atom
Boltzmann factor
41.7: Lasers
Boltzmann, Ludwig
38.1: Entropy
Boltzmann's constant
36.3: Ideal Gas Law
38.1: Entropy
39.1: Blackbody Radiation
bond (chemical)
42.1: Molecular Bonds
bond (chemical), covalent
42.1: Molecular Bonds
bond (chemical), hydrogen
42.1: Molecular Bonds
bond (chemical), ionic
42.1: Molecular Bonds
bond (chemical), metallic
42.1: Molecular Bonds
bond (chemical), van der Waals
42.1: Molecular Bonds
Bose–Einstein condensate
44.1: Elementary Particles
boson
44.1: Elementary Particles
Boyle's law
36.3: Ideal Gas Law
Bragg's law
34.5: Diffraction of X-Rays by Crystals
Brewster's angle
31.6: Polarization by Reflection
brittle
12.3: Elasticity
Broughton Suspension Bridge
28.9: Forced Oscillations
bulk modulus
12.3: Elasticity
16.2: Speed of Sound
buoyancy, neutral
14.4: Buoyancy and Archimedes' Principle
buoyant force
14.4: Buoyancy and Archimedes' Principle

C

calculus
D: Motivation of Calculus for Physics Students
calorie
37.1: Heat
Calorie (food)
37.1: Heat
calorimetry
37.1: Heat
capacitance
19.1: Capacitance
capacitance, equivalent
19.4: Capacitors in Parallel and in Series
capacitance, of an isolated sphere
19.1: Capacitance
capacitive reactance
26.3: Capacitors in an AC Circuit
capacitor
19.1: Capacitance
capacitor, cylindrical
19.1: Capacitance
capacitor, energy stored in
19.2: Energy Stored in an Electric Field
capacitor, film
19.3: Dielectrics
capacitor, parallel-plate
19.1: Capacitance
capacitor, spherical
19.1: Capacitance
capacitors, in parallel
19.4: Capacitors in Parallel and in Series
capacitors, in series
19.4: Capacitors in Parallel and in Series
Carnot cycle
38.3: Heat Engines
Carnot efficiency
38.3: Heat Engines
Carnot, Sadi
38.3: Heat Engines
cathode rays
22.3: Applications of Magnetic Fields
cavity in a conductor
17.5: Conductors in Electrostatic Equilibrium
center of curvature
32.2: Spherical Mirrors
center of gravity
12.2: Center of Gravity and Stability
center of mass
12.2: Center of Gravity and Stability
9.5: Center of Mass
centripetal acceleration
10.1: Rotational Kinematics
4.3: Uniform Circular Motion
6.3: Uniform Circular Motion
centripetal force
6.3: Uniform Circular Motion
chain reaction
43.4: Nuclear Reactions
chain rule
D.1: The Derivative: Slope as a Rate of Change
charge density
15.4: Electric Forces on Continuous Charge Distributions
charge density, linear
15.4: Electric Forces on Continuous Charge Distributions
charge density, surface
15.4: Electric Forces on Continuous Charge Distributions
charge density, volume
15.4: Electric Forces on Continuous Charge Distributions
charge distribution, continuous
15.4: Electric Forces on Continuous Charge Distributions
charge-to-mass ratio
16.4: A Point Charge in an Electric Field
charge, conservation of
15.1: Electric Charge
charge, electric
15.1: Electric Charge
charge, positive and negative
15.1: Electric Charge
charge, quantization of
15.1: Electric Charge
charging, by conduction
15.1: Electric Charge
charging, by friction
15.1: Electric Charge
charging, by induction
15.1: Electric Charge
Charles's law
36.3: Ideal Gas Law
circuit diagram
21.1: Circuit Elements and Diagrams
circular motion, uniform
4.3: Uniform Circular Motion
6: The Laws of Motion II
closed surface
17.2: Electric Flux
coaxial cable
17.E: Gauss's Law (Problems)
19.1: Capacitance
coefficient of linear expansion
36.2: Thermal Expansion
coefficient of performance
38.3: Heat Engines
coefficient of volume expansion
36.2: Thermal Expansion
coherence
33.1: Young's Double-Slit Experiment
coherent light
33.1: Young's Double-Slit Experiment
collision
9.4: Collisions
collision, elastic
9.4: Collisions
collision, inelastic
9.4: Collisions
collision, perfectly inelastic
9.4: Collisions
color charge
44.2: Basic Forces and Messenger Particles
commutator
22.3: Applications of Magnetic Fields
component vector
3.2: Components of a Vector
components, vector
3.2: Components of a Vector
C.6: Breaking a Vector into Components
compression (sound)
16.1: Sound Waves
Compton scattering
39.4: Compton Shift

Compton shift
39.4: Compton Shift

Compton wavelength
39.4: Compton Shift

Compton, Arthur
39.4: Compton Shift

conduction
37.4: Heat Transfer Mechanisms

conduction band
42.3: Electrical Conduction in Metals and Insulators

conductivity (electrical)
20.2: Resistance

conductor
15.1: Electric Charge

conductor, as an equipotential
18.2: Equipotential Surfaces

conductor, band picture
42.3: Electrical Conduction in Metals and Insulators

conductor, field at surface of
17.5: Conductors in Electrostatic Equilibrium

conductor, in electrostatic equilibrium
17.5: Conductors in Electrostatic Equilibrium

conservation of charge
15.1: Electric Charge

conservation, of angular momentum
11.S: Rotation II (Summary)

conservation, of linear momentum
9.3: Conservation of Linear Momentum

conservative force
18.1: Electric Potential
8.4: Conservative and Nonconservative Forces

continuity equation
14.5: Fluid Dynamics and Bernoulli's Principle

continuous charge distribution
15.4: Electric Forces on Continuous Charge Distributions

control rod
43.4: Nuclear Reactions

convection
37.4: Heat Transfer Mechanisms

conversion factor
1.4: Unit Conversions

cornea
32.5: Optical Instruments

corona discharge
17.5: Conductors in Electrostatic Equilibrium
18.2: Equipotential Surfaces

correspondence principle
39.1: Blackbody Radiation

cosmic microwave background
44.4: Cosmology

cosmology
44.4: Cosmology

coulomb (unit)
15.1: Electric Charge

Coulomb barrier
43.4: Nuclear Reactions

Coulomb constant
15.2: Coulomb's Law

Coulomb's law
15.2: Coulomb's Law

Coulomb's law, from Gauss's law
17.4: Applications of Gauss's Law

covalent bond
42.1: Molecular Bonds

critical angle
31.5: Total Internal Reflection

criticality
43.4: Nuclear Reactions

cross product
10.4: Torque and the Right-Hand Rule
3.5: Vector Multiplication

crystal lattice
34.5: Diffraction of X-Rays by Crystals

curie
43.3: Radioactivity

Curie temperature
23.7: Magnetic Properties of Matter

Curie's law
23.7: Magnetic Properties of Matter

current density
20.1: Electric Current

current, electric
20.1: Electric Current

curve of binding energy
43.2: Nuclear Binding Energy

cutoff frequency
39.2: Photoelectric Effect

cutoff wavelength
39.2: Photoelectric Effect

cyclic process
37.2: The First Law of Thermodynamics
38.3: Heat Engines

cyclotron
22.3: Applications of Magnetic Fields

cyclotron frequency
22.2: Motion of a Charged Particle in Magnetic Fields

cylindrical shell of charge
17.4: Applications of Gauss's Law

D

Dalton, John
41.1: Models of the Atom

damping
28.8: Damped Oscillations

damping, constant
28.8: Damped Oscillations

damping, critically damped
28.8: Damped Oscillations

damping, overdamped
28.8: Damped Oscillations

damping, underdamped
28.8: Damped Oscillations

dark energy
44.4: Cosmology

dark matter
44.4: Cosmology

daughter nucleus
43.4: Nuclear Reactions

de Broglie wavelength
39.5: de Broglie Wavelength

de Broglie, Louis
39.5: de Broglie Wavelength

decay constant
43.3: Radioactivity

decibel
30.4: Intensity and Decibels

degeneracy
40.5: Two- and Three-Dimensional Wells

degree of freedom
36.4: Kinetic Theory of Gases

density
14: Fluids

density, linear mass
29.3: Speed of a Wave

depletion zone
42.4: Semiconductors

derivative
D.1: The Derivative: Slope as a Rate of Change

diamagnetism
23.7: Magnetic Properties of Matter

dielectric
19.3: Dielectrics

dielectric constant
19.3: Dielectrics

dielectric strength
19.3: Dielectrics

dielectric, polarization of
19.3: Dielectrics

diffraction
31.1: Light as a Wave
34: Diffraction

diffraction grating
34.4: Diffraction Gratings

diffraction, circular aperture
34.2: Diffraction by a Circular Aperture

diffraction, double-slit
34.3: Diffraction by a Double Slit

diffraction, grating
34.4: Diffraction Gratings

diffraction, of X-rays
34.5: Diffraction of X-Rays by Crystals

diffraction, single-slit
34.1: Single-Slit Diffraction

dimensional analysis
1.S: Measurement (Summary)

dimensions
1.5: Dimensional Analysis and Estimation

diode
20.3: Ohm's Law
42.4: Semiconductors

diode, p–n junction
42.4: Semiconductors

diopter
32.5: Optical Instruments

dipole moment, electric
16.2: The Electric Field of Point Charges

dipole moment, magnetic
22.3: Applications of Magnetic Fields

dipole, electric
16.2: The Electric Field of Point Charges

dipole, field of
16.2: The Electric Field of Point Charges

dipole, in a nonuniform field
16.5: A Dipole in an Electric Field

dipole, magnetic
22.3: Applications of Magnetic Fields

dipole, magnetic, field of
23.3: Magnetic Dipoles

dipole, potential of
18.3: Potential of a Discrete Charge Arrangement

discriminant
C.2: Quadratic Equations
disk of charge, field of
16.3: The Electric Field of Continuous Charge Distributions
disk of charge, potential of
18.4: Potential of a Continuous Charge Arrangement
displacement
2.1: Position, Time, and Displacement
displacement current
27.1: Displacement Current
dissociation energy
42.1: Molecular Bonds
distance
2.1: Position, Time, and Displacement
DNA
34.5: Diffraction of X-Rays by Crystals
donor (semiconductor)
42.4: Semiconductors
doping
42.4: Semiconductors
Doppler effect
30.6: The Doppler Effect
Doppler effect, relativistic
35.4: Doppler Effect for Light
dot product
3.5: Vector Multiplication
7.1: Work Done by a Constant Force
double-slit experiment
33.1: Young's Double-Slit Experiment
drag coefficient
6.2: Resistive Forces
drag force
6.2: Resistive Forces
drift velocity
20.1: Electric Current
driving frequency
28.9: Forced Oscillations
dynamics
5: The Laws of Motion I

E

Earnshaw's theorem
15.3: Discrete Charge Arrangements and the Superposition Principle
echolocation
30.1: Sound Waves
eddy currents
24.2: Faraday's Law of Induction
efficiency
38.3: Heat Engines
efficiency, Carnot
38.3: Heat Engines
efficiency, thermal
38.3: Heat Engines
Einstein, Albert
35.1: History and Motivation for Special Relativity
41.7: Lasers
Einstein's postulates
35.1: History and Motivation for Special Relativity
elastic limit
12.3: Elasticity
elastic modulus
12.3: Elasticity
elastic modulus, bulk
12.3: Elasticity
elastic modulus, shear
12.3: Elasticity
elastic modulus, Young's
12.3: Elasticity
elastic potential energy
28.4: Energy in Simple Harmonic Motion
elasticity
12: Static Equilibrium and Elasticity
electric charge
15.1: Electric Charge
electric circuit
21.1: Circuit Elements and Diagrams
electric current
20.1: Electric Current
electric dipole
16.2: The Electric Field of Point Charges
electric dipole moment
16.2: The Electric Field of Point Charges
electric field
16.1: The Electric Field
electric field lines
16.1: The Electric Field
electric field, from the potential
18.5: Calculating the Field from the Potential
electric field, induced
24.2: Faraday's Law of Induction
electric field, uniform
16.3: The Electric Field of Continuous Charge Distributions
electric flux
17.2: Electric Flux
electric motor
22.3: Applications of Magnetic Fields
electric potential
18.1: Electric Potential
electric potential energy
18.1: Electric Potential
electromagnetic induction
24.2: Faraday's Law of Induction
electromagnetic spectrum
27.6: The Electromagnetic Spectrum
electromagnetic wave
27.3: Propagation of Electromagnetic Waves
electron
15.1: Electric Charge
electron affinity
42.1: Molecular Bonds
electron cloud
41.1: Models of the Atom
electron configuration
41.6: The Pauli Exclusion Principle
electron microscope
34.2: Diffraction by a Circular Aperture
39.5: de Broglie Wavelength
electron volt
18.1: Electric Potential
electron, discovery of
22.3: Applications of Magnetic Fields
electron, free
15.1: Electric Charge
electrostatic equilibrium
17.5: Conductors in Electrostatic Equilibrium
electrostatics
15.2: Coulomb's Law
electroweak force
44.2: Basic Forces and Messenger Particles
elementary charge
15.1: Electric Charge
elimination method
C.3: Solving for Two Unknowns
emf
21.1: Circuit Elements and Diagrams
emf, back
24.4: Generators and Motors
emf, induced
24.2: Faraday's Law of Induction
emf, motional
24.3: Motional EMF
emf, self-induced
25.1: Inductors
emission spectrum
41.3: The Hydrogen Atom
emissivity
37.4: Heat Transfer Mechanisms
energy
7: Work and Power
energy band
42.3: Electrical Conduction in Metals and Insulators
energy density, of a magnetic field
25.2: Energy in a Magnetic Field
energy density, of an electric field
19.2: Energy Stored in an Electric Field
energy diagram
8.5: Conservation of Mechanical Energy
energy, internal
36.4: Kinetic Theory of Gases
37.1: Heat
energy, kinetic
8.2: Kinetic Energy and the Work–Kinetic Energy Theorem
energy, mechanical
8.5: Conservation of Mechanical Energy
energy, nuclear binding
43.2: Nuclear Binding Energy
energy, potential
8.3: Potential Energy
energy, quantization of
39.1: Blackbody Radiation
energy, rest
35.5: Momentum and Energy
energy, total relativistic
35.5: Momentum and Energy
energy, zero-point
40.3: Infinite Wells
engine, heat
38.3: Heat Engines
entropy
38: Entropy and the Second Law of Thermodynamics
44.4: Cosmology
entropy, Boltzmann relation
38.1: Entropy
equations, simultaneous
C.3: Solving for Two Unknowns

equilibrium
12: Static Equilibrium and Elasticity
5.4: Equilibrium

equilibrium separation
42.1: Molecular Bonds

equilibrium, stable
8.5: Conservation of Mechanical Energy

equilibrium, static
12: Static Equilibrium and Elasticity

equilibrium, thermal
36.1: Temperature

equilibrium, unstable
8.5: Conservation of Mechanical Energy

equipartition theorem
36.4: Kinetic Theory of Gases

equipotential surface
18.2: Equipotential Surfaces

equivalence principle
35.6: General Relativity

equivalent capacitance
19.4: Capacitors in Parallel and in Series

equivalent resistance
21.3: Multiple Resistors in a Circuit: Series vs. Parallel

escape speed
13.3: Gravitational Potential Energy

estimation
1.5: Dimensional Analysis and Estimation

expansion of the universe
44.4: Cosmology

expectation value
40.1: Schrödinger's Equation and Wave Functions

eye, lens
32.5: Optical Instruments

eyepiece
32.5: Optical Instruments

F

far point
32.5: Optical Instruments

farad
19.1: Capacitance

Faraday cage
17.5: Conductors in Electrostatic Equilibrium
18.2: Equipotential Surfaces

Faraday, Michael
16: Electric Fields

Faraday's law of induction
24.2: Faraday's Law of Induction
27.2: Maxwell's Equations

farsightedness (hyperopia)
32.5: Optical Instruments

fate of the universe, Big Crunch
44.4: Cosmology

fate of the universe, Big Rip
44.4: Cosmology

fate of the universe, heat death
44.4: Cosmology

femtometer
43.1: Properties of Nuclei

Fermi energy
42.2: Free-Electron Theory of Metals

Fermi–Dirac distribution
42.2: Free-Electron Theory of Metals

fermion
44.1: Elementary Particles

ferromagnetism
23.7: Magnetic Properties of Matter

field lines
16.1: The Electric Field

field lines, magnetic
22.1: Magnetic Fields

field, electric
16.1: The Electric Field

finite well
40.4: Finite Wells

first law of thermodynamics
37.2: The First Law of Thermodynamics

fixed end
29.6: Standing Waves and Resonance

flow rate, mass
14.5: Fluid Dynamics and Bernoulli's Principle

flow rate, volume
14.5: Fluid Dynamics and Bernoulli's Principle

fluid
14.1: Fluids, Density, and Pressure

fluid, ideal
14.5: Fluid Dynamics and Bernoulli's Principle

flux, electric
17.2: Electric Flux

flux, magnetic
24.1: Magnetic Flux

focal length
32.2: Spherical Mirrors

focal point
32.2: Spherical Mirrors

force
5: The Laws of Motion I

force carrier
44.2: Basic Forces and Messenger Particles

force, buoyant
14.4: Buoyancy and Archimedes' Principle

force, centripetal
6.3: Uniform Circular Motion

force, conservative
8.4: Conservative and Nonconservative Forces

force, drag
6.2: Resistive Forces

force, net
5.1: Newton's Laws of Motion

force, non-conservative
8.4: Conservative and Nonconservative Forces

force, normal
5.2: Types of Forces

force, restoring
28.2: Hooke's Law

force, strong nuclear
43.1: Properties of Nuclei

force, tension
5.2: Types of Forces

forward bias
42.4: Semiconductors

free electron
15.1: Electric Charge

free end
29.6: Standing Waves and Resonance

free expansion
38.1: Entropy

free fall
2.6: Free Fall
4: Motion in Two Dimensions

free-body diagram
12.1: Equilibrium
14.2: Pressure as a Function of Depth
5.3: Free-Body Diagrams
6.1: Friction

free-electron model
42.2: Free-Electron Theory of Metals

free-fall acceleration
13.S: Universal Gravitation (Summary)

frequency
28.1: Simple Harmonic Motion
4.3: Uniform Circular Motion

frequency, angular
28.1: Simple Harmonic Motion

frequency, beat
30.5: Beats

frequency, driving
28.9: Forced Oscillations

frequency, fundamental
29.6: Standing Waves and Resonance

frequency, natural
28.9: Forced Oscillations

frequency, resonant
29.6: Standing Waves and Resonance

friction, coefficient of
6.1: Friction

friction, kinetic
6.1: Friction

friction, static
10.4: Torque and the Right-Hand Rule
6.1: Friction

fringe, bright
33.1: Young's Double-Slit Experiment

fringe, dark
33.1: Young's Double-Slit Experiment

fringe, spacing
33.1: Young's Double-Slit Experiment

fringing field
16.3: The Electric Field of Continuous Charge Distributions

fundamental frequency
29.6: Standing Waves and Resonance

fundamental theorem of calculus
D.3: Why the Two Ideas Undo Each Other

G

Galilean transformation
35.3: Lorentz Transformations

Galileo
35.1: History and Motivation for Special Relativity

gamma ray
43.4: Nuclear Reactions

gas, ideal
36.3: Ideal Gas Law

gas, kinetic theory
36.4: Kinetic Theory of Gases

gas, monatomic
36.4: Kinetic Theory of Gases

gauge pressure
14.S: Fluids (Summary)

gauss (unit)
22.1: Magnetic Fields

Gauss, Carl Friedrich
17.3: Gauss's Law
Gauss's law
17.3: Gauss's Law
Gauss's law, for electric fields
17.3: Gauss's Law
Gauss's law, for magnetic fields
23.6: Gauss's Law in Magnetism
27.2: Maxwell's Equations
Gaussian surface
17.3: Gauss's Law
Gay-Lussac's law
36.3: Ideal Gas Law
general relativity
35.6: General Relativity
generator, electric
24.4: Generators and Motors
geometric optics
31.2: Ray Optics Approximation
geosynchronous orbit
13.2: Applications of Universal Gravitation
gluon
44.2: Basic Forces and Messenger Particles
gold-foil experiment
41.1: Models of the Atom
gradient
18.5: Calculating the Field from the Potential
Grand Unified Theory
44.3: The Standard Model
gravitation, universal law of
13.S: Universal Gravitation (Summary)
gravitational constant
13.1: Newton's Law of Universal Gravitation
gravitational field
8.3: Potential Energy
gravitational lens
35.6: General Relativity
44.4: Cosmology
gravitational wave
33.4: Interferometry
35.6: General Relativity
ground (zero of potential)
18.1: Electric Potential
ground state
41.3: The Hydrogen Atom
grounding
15.1: Electric Charge

H

hadron
44.1: Elementary Particles
hadron, baryon
44.1: Elementary Particles
hadron, meson
44.1: Elementary Particles
half-life
43.3: Radioactivity
half-life, of an RC circuit
21.5: RC Circuits
Hall effect
22.3: Applications of Magnetic Fields
Hall voltage
22.3: Applications of Magnetic Fields
harmonic
29.6: Standing Waves and Resonance
heat
37.1: Heat
heat capacity
37.1: Heat
heat engine
38.3: Heat Engines
heat pump
38.3: Heat Engines
heat transfer, conduction
37.4: Heat Transfer Mechanisms
heat transfer, convection
37.4: Heat Transfer Mechanisms
heat transfer, radiation
37.4: Heat Transfer Mechanisms
heat, latent
37.1: Heat
heat, specific
37.1: Heat
Heisenberg uncertainty principle
39.6: Heisenberg Uncertainty Principle
Heisenberg, Werner
39.6: Heisenberg Uncertainty Principle
helical motion
22.2: Motion of a Charged Particle in Magnetic Fields
helium
41: Atomic Physics
henry (unit)
25.1: Inductors
Higgs boson
44.2: Basic Forces and Messenger Particles
Higgs field
44.2: Basic Forces and Messenger Particles
hole (semiconductor)
42.4: Semiconductors
Hooke's law
28.2: Hooke's Law
7.3: Work Done by a General Force
8.4: Conservative and Nonconservative Forces
Hubble constant
44.4: Cosmology
Hubble time
44.4: Cosmology
Hubble, Edwin
35.4: Doppler Effect for Light
Hubble's law
44.4: Cosmology
Huygens' principle
31.1: Light as a Wave
34.1: Single-Slit Diffraction
hydraulic press
14.3: Pascal's Principle
hydrogen bond
42.1: Molecular Bonds
hypotenuse
C.5: Right-Triangle Trigonometry
hysteresis
23.7: Magnetic Properties of Matter

I

ideal fluid
14.5: Fluid Dynamics and Bernoulli's Principle
ideal gas law
36.3: Ideal Gas Law
idealization
1.1: What is Physics?
image
32.1: Images and Plane Mirrors
image, real
32.1: Images and Plane Mirrors
image, virtual
32.1: Images and Plane Mirrors
impedance
26.5: Series RLC Circuits
impedance matching
21.4: Ideal vs. Real Batteries
impedance of free space
27.4: Energy Carried by Electromagnetic Waves
impulse
9.2: Impulse
impulse–momentum theorem
9.2: Impulse
index of refraction
31.4: Reflection and Refraction
32.3: Thin Lenses
33.3: Interference from Thin Films
induced charge
19.3: Dielectrics
induced electric field
24.2: Faraday's Law of Induction
induced emf
24.2: Faraday's Law of Induction
inductance
25.1: Inductors
inductance, mutual
25.1: Inductors
induction, charging by
15.1: Electric Charge
induction, electromagnetic
24.2: Faraday's Law of Induction
inductive reactance
26.4: Inductors in an AC Circuit
inductor
25.1: Inductors
inertia
5.1: Newton's Laws of Motion
inertial confinement
43.4: Nuclear Reactions
infinite potential well
40.3: Infinite Wells
infinitesimal
D.2: The Integral: Adding Up Small Pieces
inflation (cosmic)
44.4: Cosmology
instantaneous velocity
2.S: Motion in One Dimension (Summary)
insulator
15.1: Electric Charge
insulator, band picture
42.3: Electrical Conduction in Metals and Insulators
integral
D.2: The Integral: Adding Up Small Pieces
integrand
D.2: The Integral: Adding Up Small Pieces
integrated circuit
42.4: Semiconductors

intensity
27.4: Energy Carried by Electromagnetic Waves
30.4: Intensity and Decibels
intensity, of an electromagnetic wave
27.4: Energy Carried by Electromagnetic Waves
interference
33: Interference
34: Diffraction
interference, constructive
29.5: Phasors
33.1: Young's Double-Slit Experiment
interference, destructive
29.5: Phasors
33.1: Young's Double-Slit Experiment
interference, double-slit
33.1: Young's Double-Slit Experiment
interference, path-length difference
30.3: Wave Interference
interference, thin film
33.3: Interference from Thin Films
interferometer, Michelson
33.4: Interferometry
internal energy
36.4: Kinetic Theory of Gases
37.1: Heat
internal resistance
21.4: Ideal vs. Real Batteries
invariant mass
44.3: The Standard Model
inverse trigonometric functions
C.5: Right-Triangle Trigonometry
inverse-square law
13.1: Newton's Law of Universal Gravitation
27.4: Energy Carried by Electromagnetic Waves
30.4: Intensity and Decibels
ionic bond
42.1: Molecular Bonds
irreversible process
38.1: Entropy
isolated system
37.2: The First Law of Thermodynamics
isotherm
36.3: Ideal Gas Law
isotope
43.1: Properties of Nuclei

J

joule
7.1: Work Done by a Constant Force
8: Conservation of Energy
Joule heating
20.4: Electrical Power
Joule, James Prescott
37.1: Heat
junction (in a circuit)
21.1: Circuit Elements and Diagrams
junction rule
21.2: Kirchhoff's Laws and the Basic Circuit

K

kelvin
36.1: Temperature
Kepler's laws
13.S: Universal Gravitation (Summary)
Kepler's laws, first
13.4: Kepler's Laws of Planetary Motion
Kepler's laws, second
13.4: Kepler's Laws of Planetary Motion
Kepler's laws, third
13.4: Kepler's Laws of Planetary Motion
kilogram
1.S: Measurement (Summary)
kilogram, history of
1.2: Measuring Standards
kilowatt-hour
20.4: Electrical Power
kinematic equations
2.S: Motion in One Dimension (Summary)
4: Motion in Two Dimensions
kinematic equations, rotational
10.S: Rotation I (Summary)
kinematics
2: Motion in One Dimension
kinetic energy
10: Rotation I
11: Rotation II
28.4: Energy in Simple Harmonic Motion
8.1: Where Does the Work Go?
8.2: Kinetic Energy and the Work–Kinetic Energy Theorem
8.4: Conservative and Nonconservative Forces
kinetic energy, relativistic
35.5: Momentum and Energy
kinetic energy, rotational
11.1: Rotational Kinetic Energy
kinetic friction
6.1: Friction
kinetic theory of gases
36.4: Kinetic Theory of Gases
Kirchhoff's rules
21.2: Kirchhoff's Laws and the Basic Circuit
Kirchhoff's rules, junction rule
21.2: Kirchhoff's Laws and the Basic Circuit
Kirchhoff's rules, loop rule
21.2: Kirchhoff's Laws and the Basic Circuit

L

laser
41.7: Lasers
laser, diode
42.4: Semiconductors
laser, helium–neon
41.7: Lasers
latent heat
37.1: Heat
latent heat, of fusion
37.1: Heat
latent heat, of sublimation
37.1: Heat
latent heat, of vaporization
37.1: Heat
law of reflection
31.4: Reflection and Refraction
LC circuit
26.1: LC Oscillations
length contraction
35.2: Consequences of Special Relativity
length, proper
35.2: Consequences of Special Relativity
lens
32.3: Thin Lenses
lens, converging
32.3: Thin Lenses
lens, diverging
32.3: Thin Lenses
lens, thin
32.3: Thin Lenses
lensmaker's equation
32.3: Thin Lenses
Lenz's law
24.2: Faraday's Law of Induction
lepton
44.1: Elementary Particles
lepton number
44.1: Elementary Particles
lever arm
10.4: Torque and the Right-Hand Rule
light clock
35.2: Consequences of Special Relativity
light-emitting diode (LED)
42.4: Semiconductors
lightning rod
17.5: Conductors in Electrostatic Equilibrium
18.2: Equipotential Surfaces
LIGO
33.4: Interferometry
limiting case
1.1: What is Physics?
limits of integration
D.2: The Integral: Adding Up Small Pieces
line charge, field of
16.3: The Electric Field of Continuous Charge Distributions
line of charge, infinite
16.3: The Electric Field of Continuous Charge Distributions
17.4: Applications of Gauss's Law
line of charge, potential difference near
18.4: Potential of a Continuous Charge Arrangement
line of charge, potential of a finite
18.4: Potential of a Continuous Charge Arrangement
linear mass density
29.3: Speed of a Wave
D.4: Integration in Physics: Slice, Label, Add
linear momentum
9.1: Linear Momentum
Lockyer, Norman
41: Atomic Physics
longitudinal wave
29.1: Wave Motion
loop rule
21.2: Kirchhoff's Laws and the Basic Circuit
Lorentz factor
35.2: Consequences of Special Relativity
Lorentz force
22.2: Motion of a Charged Particle in Magnetic Fields
Lorentz transformation
35.3: Lorentz Transformations
luminiferous aether
35.1: History and Motivation for Special Relativity

M

macrostate
38.1: Entropy

magnetic braking
24.2: Faraday's Law of Induction

magnetic confinement
43.4: Nuclear Reactions

magnetic dipole moment
22.3: Applications of Magnetic Fields

magnetic dipole, field of
23.3: Magnetic Dipoles

magnetic domain
23.7: Magnetic Properties of Matter

magnetic field
22.1: Magnetic Fields

magnetic field lines
22.1: Magnetic Fields

magnetic field, of a current loop
23.1: Magnetic Field Due to a Current

magnetic field, of a long straight wire
23.1: Magnetic Field Due to a Current

magnetic flux
24.1: Magnetic Flux

magnetic monopole
22.1: Magnetic Fields
23.6: Gauss's Law in Magnetism
27.2: Maxwell's Equations

magnetic pole
22.1: Magnetic Fields

magnetic susceptibility
23.7: Magnetic Properties of Matter

magnetization
23.7: Magnetic Properties of Matter

magnification
32.2: Spherical Mirrors

magnification, angular
32.5: Optical Instruments

magnification, total
32.4: Optical Systems

magnitude
3.1: Scalars and Vectors

Malus's law
31.3: Polarization

mass
5.2: Types of Forces

mass defect
43.2: Nuclear Binding Energy

mass number
43.1: Properties of Nuclei

mass spectrometer
22.2: Motion of a Charged Particle in Magnetic Fields

mass–energy equivalence
35.5: Momentum and Energy

matter waves
39.5: de Broglie Wavelength

Maxwell speed distribution
36.4: Kinetic Theory of Gases

Maxwell, James Clerk
27.3: Propagation of Electromagnetic Waves

Maxwell's equations
27.2: Maxwell's Equations

mean free path
36.4: Kinetic Theory of Gases

mean free time
20.2: Resistance

mechanical energy
13.3: Gravitational Potential Energy
8.5: Conservation of Mechanical Energy

mechanical equivalent of heat
37.1: Heat

mechanical wave
29.1: Wave Motion

medium (wave)
29.1: Wave Motion

meson
44.1: Elementary Particles

metallic bond
42.1: Molecular Bonds

metastable state
41.7: Lasers

meter
1.S: Measurement (Summary)

meter, history of
1.2: Measuring Standards

Mexico City earthquake
28: Oscillations

Michelson interferometer
33.4: Interferometry

Michelson–Morley experiment
35.1: History and Motivation for Special Relativity

microgravity
13.2: Applications of Universal Gravitation

microscope
32.4: Optical Systems

microscope, compound
32.5: Optical Instruments

microstate
38.1: Entropy

microwave oven
16.5: A Dipole in an Electric Field

Millennium Bridge
28.9: Forced Oscillations

Millikan oil-drop experiment
16.4: A Point Charge in an Electric Field

mirror equation
32.2: Spherical Mirrors

mirror, concave
32.2: Spherical Mirrors

mirror, convex
32.2: Spherical Mirrors

mirror, plane
32.1: Images and Plane Mirrors

mirror, spherical
32.2: Spherical Mirrors

missing momentum
44.3: The Standard Model

missing order
34.3: Diffraction by a Double Slit

moderator (reactor)
43.4: Nuclear Reactions

molar mass
36.3: Ideal Gas Law

molar specific heat
37.3: Thermodynamic Processes

mole
36.3: Ideal Gas Law

moment arm
10.4: Torque and the Right-Hand Rule

moment of inertia
10.3: Rotational Inertia
10.4: Torque and the Right-Hand Rule

momentum, angular
11: Rotation II

momentum, linear
9.1: Linear Momentum

momentum, of a photon
39.3: Photons

momentum, relativistic
35.5: Momentum and Energy

Moore's law
42.4: Semiconductors

MOSFET
42.4: Semiconductors

motion, bound
8.5: Conservation of Mechanical Energy

motion, unbound
8.5: Conservation of Mechanical Energy

motional emf
24.3: Motional EMF

multiplicity
38.1: Entropy

mutual inductance
25.1: Inductors

N

natural frequency
28.9: Forced Oscillations

near point
32.5: Optical Instruments

nearsightedness (myopia)
32.5: Optical Instruments

net force
5.1: Newton's Laws of Motion

neutral, electrically
15.1: Electric Charge

neutrino
43.4: Nuclear Reactions

neutron
15.1: Electric Charge

neutron number
43.1: Properties of Nuclei

newton (unit)
5.1: Newton's Laws of Motion

Newton's law of universal gravitation
13.S: Universal Gravitation (Summary)

Newton's laws, first
5.S: The Laws of Motion I (Summary)

Newton's laws, second
10: Rotation I
5.S: The Laws of Motion I (Summary)
6: The Laws of Motion II

Newton's laws, second, for rotation
10.4: Torque and the Right-Hand Rule

Newton's laws, third
13.1: Newton's Law of Universal Gravitation
14.5: Fluid Dynamics and Bernoulli's Principle
5.S: The Laws of Motion I (Summary)

node
29.6: Standing Waves and Resonance

non-conservative force
8.4: Conservative and Nonconservative Forces

normal (optics)
31.4: Reflection and Refraction

normal force
5.2: Types of Forces
6.1: Friction

normalization
40.1: Schrödinger's Equation and Wave Functions

nuclear fission
43.4: Nuclear Reactions

nuclear fusion
40.4: Finite Wells
43.4: Nuclear Reactions

nuclear reaction
43.4: Nuclear Reactions

nucleon
43.1: Properties of Nuclei

nucleus
41.1: Models of the Atom
43.1: Properties of Nuclei

nuclide
43.1: Properties of Nuclei

O

objective lens
32.5: Optical Instruments

ohm (unit)
20.2: Resistance

Ohm's law
20.3: Ohm's Law

ohmic device
20.3: Ohm's Law

open circuit
21.1: Circuit Elements and Diagrams

optical cavity
41.7: Lasers

optical fiber
31.5: Total Internal Reflection

optical system
32.4: Optical Systems

optical tweezers
27.5: Momentum and Radiation Pressure of Electromagnetic Waves

optics
31: Introduction to Optics

order (interference)
33.1: Young's Double-Slit Experiment

order (interference), missing
34.3: Diffraction by a Double Slit

order-of-magnitude estimate
1.5: Dimensional Analysis and Estimation

origin
2.1: Position, Time, and Displacement

P

p–n junction
42.4: Semiconductors

P–V diagram
36.3: Ideal Gas Law

pair production
44.1: Elementary Particles

parabola
4.2: Projectile Motion

parallel plates, field between
16.3: The Electric Field of Continuous Charge Distributions

parallel-axis theorem
10.S: Rotation I (Summary)

paramagnetism
23.7: Magnetic Properties of Matter

parent nucleus
43.4: Nuclear Reactions

pascal
14.1: Fluids, Density, and Pressure

Pascal's principle
14.3: Pascal's Principle

path-length difference
30.3: Wave Interference
33.1: Young's Double-Slit Experiment

Pauli exclusion principle
41.6: The Pauli Exclusion Principle
44.1: Elementary Particles

Pauli, Wolfgang
41.6: The Pauli Exclusion Principle

pendulum, physical
28.7: Physical Pendulum

pendulum, simple
28.5: The Simple Pendulum

pendulum, torsion
28.6: Torsion Pendulum

penetration depth
40.4: Finite Wells

perihelion
13.4: Kepler's Laws of Planetary Motion

period
28.1: Simple Harmonic Motion
4.3: Uniform Circular Motion

periodic motion
28.1: Simple Harmonic Motion

periodic table
41.6: The Pauli Exclusion Principle

periodic wave (traveling)
29.1: Wave Motion

permanent set
12.3: Elasticity

permeability of free space
23.1: Magnetic Field Due to a Current
27.3: Propagation of Electromagnetic Waves

permittivity
19.3: Dielectrics

permittivity of free space
15.2: Coulomb's Law
27.3: Propagation of Electromagnetic Waves

phase change
37.1: Heat

phase constant
28.3: Mass–Spring System
29.2: The Wave Equation

phase constant, of an ac circuit
26.5: Series RLC Circuits

phase difference
30.3: Wave Interference
33.2: Interference and Double-Slit Intensity

phasor
29.5: Phasors
33.2: Interference and Double-Slit Intensity
34.1: Single-Slit Diffraction
C.8: Sinusoids and Phasors

phasor, in ac circuits
26.2: Resistors in an AC Circuit

photoelectric effect
39.2: Photoelectric Effect

photon
27.5: Momentum and Radiation Pressure of Electromagnetic Waves
39.3: Photons

physical pendulum
28.7: Physical Pendulum

physics
1.1: What is Physics?

pillbox (Gaussian surface)
17.4: Applications of Gauss's Law

Planck, Max
39.1: Blackbody Radiation

Planck's constant
39.1: Blackbody Radiation

Planck's constant, reduced
39.6: Heisenberg Uncertainty Principle

Planck's radiation law
39.1: Blackbody Radiation

plane mirror
32.1: Images and Plane Mirrors

plane of charge, infinite
16.3: The Electric Field of Continuous Charge Distributions
17.4: Applications of Gauss's Law

plane wave
31.1: Light as a Wave

point charge
15.2: Coulomb's Law

polarization
31.3: Polarization

polarization, by reflection
31.6: Polarization by Reflection

polarization, filter
31.3: Polarization

polarization, of a dielectric
19.3: Dielectrics

polarization, of charge
15.1: Electric Charge

polarization, unpolarized light
31.3: Polarization

polarizing filter
31.3: Polarization

population inversion
41.7: Lasers

position
2.1: Position, Time, and Displacement

position vector
4.1: Kinematics in Two Dimensions

potential barrier
40.4: Finite Wells

potential difference
18.1: Electric Potential

potential energy
8.1: Where Does the Work Go?
8.3: Potential Energy

potential energy, elastic
28.4: Energy in Simple Harmonic Motion
8.3: Potential Energy

potential energy, electric
18.1: Electric Potential

potential energy, gravitational
8.3: Potential Energy
8.4: Conservative and Nonconservative Forces
potential energy, gravitational, general form
13.S: Universal Gravitation (Summary)
potential energy, of a magnetic dipole
22.3: Applications of Magnetic Fields
potential energy, of a system of charges
18.3: Potential of a Discrete Charge Arrangement
potential energy, of an electric dipole
16.5: A Dipole in an Electric Field
potential well, finite
40.4: Finite Wells
potential well, infinite
40.3: Infinite Wells
potential, electric
18.1: Electric Potential
potential, of a point charge
18.3: Potential of a Discrete Charge Arrangement
power
7.4: Power
power factor
26.6: Power in AC Circuits
power, average
7.4: Power
power, electrical
20.4: Electrical Power
power, instantaneous
7.4: Power
power, of a lens
32.5: Optical Instruments
Poynting vector
27.4: Energy Carried by Electromagnetic Waves
precession
11.5: Precession
precision
1.S: Measurement (Summary)
prefixes, SI
1.3: Scientific Notation and Significant Figures
presbyopia
32.5: Optical Instruments
pressure
14.1: Fluids, Density, and Pressure
pressure, absolute
14.2: Pressure as a Function of Depth
pressure, atmospheric
14.2: Pressure as a Function of Depth
pressure, gauge
14.S: Fluids (Summary)
pressure, radiation
27.5: Momentum and Radiation Pressure of Electromagnetic Waves
pressure, variation in sound
30.1: Sound Waves
primary mirror
32.5: Optical Instruments
principal axis
32.2: Spherical Mirrors
principal rays
32.2: Spherical Mirrors
principle of relativity
35.1: History and Motivation for Special Relativity
probability density
40.1: Schrödinger's Equation and Wave Functions
probability density, radial
41.5: The Quantum Model of the Hydrogen Atom
product, cross (vector)
3.5: Vector Multiplication
product, dot (scalar)
3.5: Vector Multiplication
projectile motion
4.S: Motion in Two Dimensions (Summary)
projectile motion, range
4.2: Projectile Motion
projection, scalar
3.5: Vector Multiplication
projection, vector
3.5: Vector Multiplication
proper length
35.2: Consequences of Special Relativity
proper time
35.2: Consequences of Special Relativity
proportional limit
12.3: Elasticity
proton
15.1: Electric Charge
proton–proton chain
43.4: Nuclear Reactions
pulley
5.2: Types of Forces
pulse wave
29.1: Wave Motion
pumping (laser)
41.7: Lasers
Pythagorean theorem
C.5: Right-Triangle Trigonometry

Q

Q value
43.4: Nuclear Reactions
quadratic equation
C.2: Quadratic Equations
quadratic formula
C.2: Quadratic Equations
quantization
39.1: Blackbody Radiation
40.2: Energies of Trapped Electrons
quantum chromodynamics
44.2: Basic Forces and Messenger Particles
quantum dot
40: A Deeper Look at Quantum Mechanics
quantum electrodynamics
44.2: Basic Forces and Messenger Particles
quantum number
40.2: Energies of Trapped Electrons
41.5: The Quantum Model of the Hydrogen Atom
quantum number, orbital
41.5: The Quantum Model of the Hydrogen Atom
quantum number, orbital magnetic
41.5: The Quantum Model of the Hydrogen Atom
quantum number, principal
41.5: The Quantum Model of the Hydrogen Atom
quantum number, spin magnetic
41.6: The Pauli Exclusion Principle
quantum tunneling
40.4: Finite Wells
quark
44.1: Elementary Particles

R

radial probability density
41.5: The Quantum Model of the Hydrogen Atom
radian
10.1: Rotational Kinematics
radiation (thermal)
37.4: Heat Transfer Mechanisms
radiation pressure
27.5: Momentum and Radiation Pressure of Electromagnetic Waves
radiation, from accelerating charges
27.6: The Electromagnetic Spectrum
radioactive dating
43.3: Radioactivity
radioactive decay
43.3: Radioactivity
radioactive decay, alpha
43.4: Nuclear Reactions
radioactive decay, beta
43.4: Nuclear Reactions
radioactive decay, gamma
43.4: Nuclear Reactions
radioactivity
43.3: Radioactivity
radiocarbon dating
43.3: Radioactivity
range (projectile)
4.2: Projectile Motion
rarefaction
30.1: Sound Waves
rate of change
D.1: The Derivative: Slope as a Rate of Change
ray
31.2: Ray Optics Approximation
ray approximation
31.2: Ray Optics Approximation
ray diagram
32.2: Spherical Mirrors
ray optics
31.2: Ray Optics Approximation
Rayleigh–Jeans law
39.1: Blackbody Radiation
Rayleigh's criterion
34.2: Diffraction by a Circular Aperture
RC circuit
21.5: RC Circuits
RC circuit, time constant
21.5: RC Circuits
reactance, capacitive
26.3: Capacitors in an AC Circuit
reactance, inductive
26.4: Inductors in an AC Circuit
reaction energy
43.4: Nuclear Reactions

real image
32.1: Images and Plane Mirrors
recoil
9.3: Conservation of Linear Momentum
recombination (cosmic)
44.4: Cosmology
redshift
35.4: Doppler Effect for Light
reduced mass
42.1: Molecular Bonds
reduced Planck constant
39.6: Heisenberg Uncertainty Principle
reference frame
4.4: Relative Motion
reference frame, inertial
35: Relativity
5.1: Newton's Laws of Motion
reference intensity
30.4: Intensity and Decibels
reflection coefficient
40.4: Finite Wells
reflection, diffuse
31.4: Reflection and Refraction
reflection, law of
31.4: Reflection and Refraction
32.1: Images and Plane Mirrors
reflection, of a wave
29.6: Standing Waves and Resonance
reflection, phase shift on
33.3: Interference from Thin Films
reflection, specular
31.4: Reflection and Refraction
reflection, total internal
31.5: Total Internal Reflection
refraction
31.4: Reflection and Refraction
32.3: Thin Lenses
refraction, index of
31.4: Reflection and Refraction
refrigerator
38.3: Heat Engines
relative motion
4.S: Motion in Two Dimensions (Summary)
relativistic velocity addition
35.2: Consequences of Special Relativity
relativity of simultaneity
35.3: Lorentz Transformations
relativity, general
35.6: General Relativity
relativity, special
35: Relativity
resistance
20.2: Resistance
resistance, equivalent
21.3: Multiple Resistors in a Circuit: Series vs. Parallel
resistivity
20.2: Resistance
resistor
20.2: Resistance
resistors, in parallel
21.3: Multiple Resistors in a Circuit: Series vs. Parallel
resistors, in series
21.3: Multiple Resistors in a Circuit: Series vs. Parallel
resolution (optical)
34.2: Diffraction by a Circular Aperture
resolution (optical), Rayleigh's criterion
34.2: Diffraction by a Circular Aperture
resolution of vectors
3.S: Vectors (Summary)
resonance
28: Oscillations
28.9: Forced Oscillations
29.6: Standing Waves and Resonance
resonance, in an RLC circuit
26.5: Series RLC Circuits
resonant frequency
29.6: Standing Waves and Resonance
resonant frequency, of an RLC circuit
26.5: Series RLC Circuits
rest energy
35.5: Momentum and Energy
restoring force
28.2: Hooke's Law
retina
32.5: Optical Instruments
reverse bias
42.4: Semiconductors
reversible process
38.1: Entropy
Riemann sum
D.2: The Integral: Adding Up Small Pieces
right triangle
C.5: Right-Triangle Trigonometry
right-hand rule
10.4: Torque and the Right-Hand Rule
3.5: Vector Multiplication
right-hand rule, for magnetic force
22.1: Magnetic Fields
right-hand rule, for the field of a current
23.1: Magnetic Field Due to a Current
ring of charge, field of
16.3: The Electric Field of Continuous Charge Distributions
ring of charge, potential of
18.4: Potential of a Continuous Charge Arrangement
RL circuit
25.3: RL Circuits
RL circuit, time constant
25.3: RL Circuits
rms value
26.2: Resistors in an AC Circuit
rocket equation
9.E: Linear Momentum and Center of Mass (Problems)
rolling motion
10.2: Rolling Motion
rolling without slipping
10.2: Rolling Motion
root-mean-square speed
36.4: Kinetic Theory of Gases
root-mean-square, current and voltage
26.2: Resistors in an AC Circuit
rotation curve, galactic
44.4: Cosmology
rotational inertia
10.3: Rotational Inertia
rotational kinematics
10.S: Rotation I (Summary)
Rutherford scattering
41.1: Models of the Atom
Rutherford, Ernest
41.1: Models of the Atom
Rydberg constant
41.3: The Hydrogen Atom
Rydberg formula
41.3: The Hydrogen Atom

S

safety factor
12.3: Elasticity
scalar
3.1: Scalars and Vectors
scalar component
3.2: Components of a Vector
scalar triple product
3.5: Vector Multiplication
scanning tunneling microscope
40.4: Finite Wells
Schrödinger equation
40.1: Schrödinger's Equation and Wave Functions
Schrödinger, Erwin
41.1: Models of the Atom
science
1.1: What is Physics?
scientific notation
1.S: Measurement (Summary)
secant line
D.1: The Derivative: Slope as a Rate of Change
second
1.S: Measurement (Summary)
second law of thermodynamics
38: Entropy and the Second Law of Thermodynamics
second, history of
1.2: Measuring Standards
secondary maxima
34.1: Single-Slit Diffraction
self-induced emf
25.1: Inductors
self-inductance
25.1: Inductors
semiconductor
15.1: Electric Charge
20.2: Resistance
semiconductor, band picture
42.3: Electrical Conduction in Metals and Insulators
semiconductor, intrinsic
42.4: Semiconductors
semiconductor, n-type
42.4: Semiconductors
semiconductor, p-type
42.4: Semiconductors
shear modulus
12.3: Elasticity
shell (atomic)
41.5: The Quantum Model of the Hydrogen Atom
shell theorem
13.S: Universal Gravitation (Summary)

shell theorem, for charge
16.3: The Electric Field of Continuous Charge Distributions
17.4: Applications of Gauss's Law
shielding, electrostatic
17.5: Conductors in Electrostatic Equilibrium
short circuit
21.1: Circuit Elements and Diagrams
SI prefixes
1.3: Scientific Notation and Significant Figures
SI units
1.2: Measuring Standards
sign convention (optics)
32.2: Spherical Mirrors
significant figures
1.S: Measurement (Summary)
simple harmonic motion
28.2: Hooke's Law
simple pendulum
28.5: The Simple Pendulum
simultaneity
35.3: Lorentz Transformations
simultaneous equations
C.3: Solving for Two Unknowns
single-slit diffraction
34.1: Single-Slit Diffraction
sinusoidal wave
29.1: Wave Motion
slit width
34.1: Single-Slit Diffraction
slope
D.1: The Derivative: Slope as a Rate of Change
small-angle approximation
28.5: The Simple Pendulum
Snell's law
31.4: Reflection and Refraction
solar cell
42.4: Semiconductors
solar constant
27.4: Energy Carried by Electromagnetic Waves
solenoid
23.5: Solenoids and Toroids
sonar
30.1: Sound Waves
sound intensity level
30.4: Intensity and Decibels
sound wave
30.1: Sound Waves
sound, intensity
30.4: Intensity and Decibels
sound, intensity level
30.4: Intensity and Decibels
sound, speed of
30.2: Speed of Sound
space quantization
41.5: The Quantum Model of the Hydrogen Atom
special relativity
35: Relativity
specific heat capacity
37.1: Heat
specific heat, molar
37.3: Thermodynamic Processes
spectral fingerprint
41.4: Atomic Spectra of Gases
spectral series
41.4: Atomic Spectra of Gases
spectral series, Balmer
41.4: Atomic Spectra of Gases
spectral series, Lyman
41.4: Atomic Spectra of Gases
spectral series, Paschen
41.4: Atomic Spectra of Gases
spectroscopy
34.4: Diffraction Gratings
41.4: Atomic Spectra of Gases
spectrum, absorption
41.3: The Hydrogen Atom
spectrum, emission
41.3: The Hydrogen Atom
speed
2.S: Motion in One Dimension (Summary)
speed distribution, Maxwell
36.4: Kinetic Theory of Gases
speed of light
27.3: Propagation of Electromagnetic Waves
speed, average
2.2: Speed and Velocity
speed, instantaneous
2.2: Speed and Velocity
speed, most probable
36.4: Kinetic Theory of Gases
speed, of a wave
29.2: The Wave Equation
speed, of sound
30.2: Speed of Sound
speed, root-mean-square
36.4: Kinetic Theory of Gases
sphere of charge, field of
16.3: The Electric Field of Continuous Charge Distributions
sphere of charge, potential of
18.4: Potential of a Continuous Charge Arrangement
spherical mirror
32.2: Spherical Mirrors
spin
41.6: The Pauli Exclusion Principle
44.1: Elementary Particles
spontaneous emission
41.7: Lasers
spring constant
28.2: Hooke's Law
7.3: Work Done by a General Force
standard (of measurement)
1.2: Measuring Standards
Standard Model
44.1: Elementary Particles
standard temperature and pressure
36.3: Ideal Gas Law
standard-position angle
3.2: Components of a Vector
standing wave
29.6: Standing Waves and Resonance
40.2: Energies of Trapped Electrons
41.2: Energies of a Trapped Electron in an Atom
state variable
37.2: The First Law of Thermodynamics
38.1: Entropy
static equilibrium
12: Static Equilibrium and Elasticity
static friction
6.1: Friction
statically indeterminate
12.1: Equilibrium
Stefan–Boltzmann constant
37.4: Heat Transfer Mechanisms
Stefan–Boltzmann law
37.4: Heat Transfer Mechanisms
stimulated emission
41.7: Lasers
stopping potential
18.1: Electric Potential
39.2: Photoelectric Effect
strain
12.3: Elasticity
strangeness
44.1: Elementary Particles
stress
12.3: Elasticity
stress, thermal
36.2: Thermal Expansion
strong nuclear force
43.1: Properties of Nuclei
44.1: Elementary Particles
subshell
41.5: The Quantum Model of the Hydrogen Atom
substitution method
C.3: Solving for Two Unknowns
supercapacitor
19.1: Capacitance
superconductor
20.2: Resistance
superposition principle
29.5: Phasors
superposition principle, for electric forces
15.3: Discrete Charge Arrangements and the Superposition Principle
switch
21.1: Circuit Elements and Diagrams
symmetry, cylindrical
17.1: The Importance of Symmetry
symmetry, planar
17.1: The Importance of Symmetry
symmetry, spherical
17.1: The Importance of Symmetry
synchrotron
22.3: Applications of Magnetic Fields
synchrotron radiation
22.3: Applications of Magnetic Fields
system
8.1: Where Does the Work Go?

T

tangent line
D.1: The Derivative: Slope as a Rate of Change
telescope
32.4: Optical Systems
telescope, reflecting
32.5: Optical Instruments
temperature coefficient of resistivity
20.2: Resistance
temperature scale, Celsius
36.1: Temperature

temperature scale, Fahrenheit
36.1: Temperature
temperature scale, Kelvin
36.1: Temperature
temperature, absolute zero
36.1: Temperature
tension
5.2: Types of Forces
terminal speed
6.2: Resistive Forces
terminal voltage
21.1: Circuit Elements and Diagrams
tesla (unit)
22.1: Magnetic Fields
test charge
16.1: The Electric Field
Theory of Everything
44.3: The Standard Model
thermal conductivity
37.4: Heat Transfer Mechanisms
thermal contact
36.1: Temperature
thermal energy
8.1: Where Does the Work Go?
thermal equilibrium
36.1: Temperature
thermal expansion, coefficient of linear
36.2: Thermal Expansion
thermal expansion, coefficient of volume
36.2: Thermal Expansion
thermal physics
36: Temperature and Kinetic Theory of Gases
thermal reservoir
38.3: Heat Engines
thermal resistance
37.4: Heat Transfer Mechanisms
thermal stress
36.2: Thermal Expansion
thermodynamic process
37.3: Thermodynamic Processes
thermodynamic process, adiabatic
37.3: Thermodynamic Processes
thermodynamic process, cyclic
38.3: Heat Engines
thermodynamic process, irreversible
38.1: Entropy
thermodynamic process, isobaric
37.3: Thermodynamic Processes
thermodynamic process, isothermal
37.3: Thermodynamic Processes
thermodynamic process, isovolumetric
37.3: Thermodynamic Processes
thermodynamic process, reversible
38.1: Entropy
thermodynamics, first law
37.2: The First Law of Thermodynamics
thermodynamics, second law
38: Entropy and the Second Law of Thermodynamics
thermodynamics, zeroth law
36.1: Temperature
thermometer
36.1: Temperature
thermonuclear fusion
43.4: Nuclear Reactions
thin film interference
33.3: Interference from Thin Films
thin lens
32.3: Thin Lenses
Thomson, J. J.
41.1: Models of the Atom
threshold energy
43.4: Nuclear Reactions
time constant
21.5: RC Circuits
time constant, inductive
25.3: RL Circuits
time dilation
35.2: Consequences of Special Relativity
time dilation, gravitational
35.6: General Relativity
time, proper
35.2: Consequences of Special Relativity
toroid
23.5: Solenoids and Toroids
torque
10.4: Torque and the Right-Hand Rule
12.1: Equilibrium
torque, on a current loop
22.3: Applications of Magnetic Fields
torque, on an electric dipole
16.5: A Dipole in an Electric Field
torsion constant
28.6: Torsion Pendulum
total internal reflection
31.5: Total Internal Reflection
trajectory
4.S: Motion in Two Dimensions (Summary)
transfer variable
37.2: The First Law of Thermodynamics
transformer
25.4: Transformers
transistor
42.4: Semiconductors
transistor, field-effect
42.4: Semiconductors
transmission coefficient
40.4: Finite Wells
transverse wave
27.3: Propagation of Electromagnetic Waves
29.1: Wave Motion
triboelectric series
15.1: Electric Charge
trigonometric functions
C.5: Right-Triangle Trigonometry
tube length
32.5: Optical Instruments
tuned mass damper
28.4: Energy in Simple Harmonic Motion
tunneling
40.4: Finite Wells
turning point
8.S: Conservation of Energy (Summary)

U

ultimate strength
12.3: Elasticity
ultrasound
30.1: Sound Waves
ultraviolet catastrophe
39.1: Blackbody Radiation
uncertainty principle
39.6: Heisenberg Uncertainty Principle
uniform circular motion
4.3: Uniform Circular Motion
uniform electric field
16.3: The Electric Field of Continuous Charge Distributions
unit conversion
1.4: Unit Conversions
unit vector
3.3: Unit Vectors
unit vector, standard basis
3.3: Unit Vectors
units, atomic mass
43.1: Properties of Nuclei
units, base
1.2: Measuring Standards
units, becquerel
43.3: Radioactivity
units, calorie
37.1: Heat
units, curie
43.3: Radioactivity
units, derived
1.2: Measuring Standards
units, femtometer
43.1: Properties of Nuclei
units, kilogram
1.S: Measurement (Summary)
units, meter
1.S: Measurement (Summary)
units, second
1.S: Measurement (Summary)
units, SI
1.2: Measuring Standards
universal gas constant
36.3: Ideal Gas Law
unpolarized light
31.3: Polarization

V

valence band
42.3: Electrical Conduction in Metals and Insulators
valence electron
41.6: The Pauli Exclusion Principle
valley of stability
43.4: Nuclear Reactions
van der Waals bond
42.1: Molecular Bonds
variable of integration
D.2: The Integral: Adding Up Small Pieces
vector
3: Vectors
vector, components
3.2: Components of a Vector
5.3: Free-Body Diagrams
velocity selector
22.2: Motion of a Charged Particle in Magnetic Fields

velocity, angular
10.1: Rotational Kinematics
velocity, average
2.S: Motion in One Dimension (Summary)
velocity, instantaneous
2.S: Motion in One Dimension (Summary)
D.1: The Derivative: Slope as a Rate of Change
velocity, relative
4.S: Motion in Two Dimensions (Summary)
velocity, relativistic addition
35.2: Consequences of Special Relativity
virtual image
32.1: Images and Plane Mirrors
virtual object
32.4: Optical Systems
virtual photon
44.2: Basic Forces and Messenger Particles
volt
18.1: Electric Potential
Volta, Alessandro
18.1: Electric Potential
voltage
18.1: Electric Potential
voltage divider
21.3: Multiple Resistors in a Circuit: Series vs. Parallel
voltmeter
21.1: Circuit Elements and Diagrams

W

watt
7.4: Power
wave equation
29.2: The Wave Equation
wave function
29.2: The Wave Equation
40.1: Schrödinger's Equation and Wave Functions
wave number
29.2: The Wave Equation
wave speed
29.2: The Wave Equation
wave, electromagnetic
27.3: Propagation of Electromagnetic Waves
wave, longitudinal
29.1: Wave Motion
wave, mechanical
29.1: Wave Motion
wave, periodic
29.1: Wave Motion
wave, plane
31.1: Light as a Wave
wave, pulse
29.1: Wave Motion
wave, sinusoidal
29.1: Wave Motion
wave, sound
30.1: Sound Waves
wave, standing
29.6: Standing Waves and Resonance
wave, transverse
29.1: Wave Motion
wavefront
31.1: Light as a Wave
wavelength
27.3: Propagation of Electromagnetic Waves
29.2: The Wave Equation
weber (unit)
24.1: Magnetic Flux
weight
5.2: Types of Forces
weight, apparent
5.5: Applying Newton's Laws
Wien's displacement law
39.1: Blackbody Radiation
work
7.1: Work Done by a Constant Force
work function
39.2: Photoelectric Effect
work–kinetic energy theorem
11.2: Rotational Work
8.2: Kinetic Energy and the Work–Kinetic Energy Theorem
work, by a variable force
7.S: Work and Power (Summary)
work, done by gravity
7.S: Work and Power (Summary)
work, net
7.1: Work Done by a Constant Force
work, path dependence
8.5: Conservation of Mechanical Energy

X

X-rays
34.5: Diffraction of X-Rays by Crystals
39.4: Compton Shift

Y

yield strength
12.3: Elasticity
Young, Thomas
33.1: Young's Double-Slit Experiment
Young's modulus
12.3: Elasticity

Z

zero-point energy
40.3: Infinite Wells
zeroth law of thermodynamics
36.1: Temperature

This page titled Index is shared under a Public Domain license and was authored, remixed, and/or curated by Michael Ragsdale.

DETAILED LICENSING

Detailed licensing information for this resource is not available at this time.

This page titled Physics: A Foundation for Scientists and Engineers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.